Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Walter V. Bingham, *Washington, D. C.*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Table of Contents

The Concentric Organization Chart: C. G. Browne 375
Custom Made Systems of Job Evaluation: J. S. Gray 378
Rating Training and Experience: L. K. Burke and E. K. Taylor 381
Measuring the Level of Abstraction: R. Flesch 384
Personality Characteristics of Embalmer Trainees: D. N. Wiener and W. Simon 391
The Guilford-Zimmerman Temperament Survey and Certain Related Personality Tests: C. Gilbert 394
The Non-Respondent Problem in Questionnaire Research: W. L. Barnette, Jr. 397
Reactions of Veterans to Counseling: W. L. Barnette, Jr. 399
The Wechsler-Bellevue Intelligence Scale and High School Achievement: A. N. Frandsen 406
The Minnesota Clerical Test: Sex Differences and Norms for College Groups: O. E. de Cillis Engelhardt 412
An IBM Card Profile for the Strong Vocational Interest Blank: W. L. Layton 415
The Super-Roper Technique as a Measure of Interest in Nursing: L. Navran 417
Acuity Differences between the Two Eyes and Job Performance: N. C. Kephart and J. M. Mason 423
Point Centering of Signals on an Area: A. Ford, D. Rigler, and G. E. Dugan 429
Influence of Friction in Making Settings on a Linear Scale: W. L. Jenkins, L. O. Maas, and D. Rigler 434
The Galvanic Skin Response as a Test of Advertising Impact: E. Golin and S. B. Lyerly .. 440
Book Reviews 444
New Books, Monographs, and Pamphlets 454

American Psychological Association

Vol. 34, No. 6 December, 1950

Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

Subscriptions and business communications should be sent to

American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.

Articles for publication and books for review should be sent to the Editor

Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described by J. E. Anderson and W. L. Valentine in "The preparation of articles for publication in the journals of the American Psychological Association," *Psychol. Bull.*, 1944, 41, 345–376. A reprint of this article will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1950, by The American Psychological Association, Inc.

Journal of Applied Psychology

Edited by

Donald G. Paterson

University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Walter V. Bingham, *Washington, D. C.*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Volume 34, 1950

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40,
P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1950, by The American Psychological Association, Inc.

LANCASTER PRESS, INC., LANCASTER, PA.

Contents of Volume 34

Articles

Ammons, R. B., Butler, M. N. and Herzig, S. A. **A Projective Test for Vocational Research and Guidance at the College Level** ... 198
Baas, M. L. **Kuder Interest Patterns of Psychologists** ... 115
Barnette, W. L., Jr. **The Non-Respondent Problem in Questionnaire Research** ... 397
Barnette, W. L., Jr. **Reactions of Veterans to Counseling** ... 399
Berdie, R. F. **Scores on the Strong Vocational Interest Blank and the Kuder Preference Record in Relation to Self Ratings** ... 42
Bittner, R. H. and Rundquist, E. A. **The Rank-Comparison Rating Method** ... 171
Brayfield, A. H. and Reed, P. A. **How Readable are Occupational Information Booklets?** ... 325
Browne, C. G. **Study of Executive Leadership in Business. II. Social Group Patterns** ... 12
Browne, C. G. **Study of Executive Leadership in Business. III. Goal and Achievement Index** ... 82
Browne, C. G. **The Concentric Organization Chart** ... 375
Burke, L. K. and Taylor, E. K. **Rating Training and Experience** ... 381
Campbell, D. T. and Mohr, P. J. **The Effect of Ordinal Position upon Responses to Items in a Check List** ... 62
Carr, E. R. and Rothe, H. F. **Validity of an Objectivity Key on a Short Industrial Personality Questionnaire** ... 178
Carter, H. L. J. **A Combined Oral Reading and Psychogalvanic Response Technique for Investigating Certain Reading Abilities of College Students** ... 267
Cook, Lt. Comdr. E. B. and Wherry, R. J. **A Factor Analysis of MMPI and Aptitude Test Data** ... 260
Corso, J. F. and Lewis, D. **Preferred Rate and Extent of the Frequency Vibrato** ... 206
Cottle, W. C. **Card Versus Booklet Forms of the MMPI** ... 255
Cover, C. B. and Pressey, S. L. **Age and Route Sales Efficiency** ... 229
Davis, M. K. and Tiffin, J. **Cross Validation of an Abbreviated Point Job Evaluation System** ... 225
de Cillis, O. E. and Orbison, W. D. **A Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI** ... 338
DuBois, P. H. and Watson, R. I. **The Selection of Patrolmen** ... 90
Dunlap, J. W. **The Effect of Color in Direct Mail Advertising** ... 280
Edwards, A. S. **The Myth of Chronological Age** ... 316
Ely, J. H., Kephart, N. C. and Tiffin, J. **Ortho-Rater Norms and Sex Differences** ... 232
Engelhardt, O. E. de Cillis. **The Minnesota Clerical Test: Sex Differences and Norms for College Groups** ... 412
England, A. O. **Getting Your Message Across by Plain Talk** ... 182
Farr, J. N. **Readability and Interest Values in an Employee Handbook** ... 16
Flesch, R. **Measuring the Level of Abstraction** ... 384
Ford, A., Rigler, D. and Dugan, G. E. **Point Centering of Signals on an Area** ... 429
Ford, A. H. **Prediction of Academic Success in Three Schools of Nursing** ... 186
Frandsen, A. N. **The Wechsler-Bellevue Intelligence Scale and High School Achievement** ... 406
Gaudet, F. J., Carli, A. R. and Dennegar, L. S. **Attitudes of Veterans toward Vocational Guidance Services** ... 347

Gilbert, C. **The Guilford-Zimmerman Temperament Survey and Certain Related Personality Tests** ... 394
Golin, E. and Lyerly, S. B. **The Galvanic Skin Response as a Test of Advertising Impact** ... 440
Goodman, C. H. **The MacQuarrie Test for Mechanical Ability. IV. Time and Motion Analysis** ... 27
Gray, J. S. **Custom Made Systems of Job Evaluation** ... 378
Gray, J. S. and Prevetta, P. **Fluorescent Light Versus Daylight** ... 235
Griffith, J. W., Kerr, W. A. and Mayo, T. B., Jr. **Changes in Subjective Fatigue and Readiness for Work During the Eight-Hour Shift** ... 163
Hay, E. N. **Cross-Validation of Clerical Aptitude Tests** ... 153
Hay, E. N. **The Application of Weber's Law to Job Evaluation Estimates** ... 102
Hayes, P. M., Jenkins, J. J. and Walker, B. J. **Reliability of the Flesch Readability Formulas** ... 22
Jenkins, W. L., Maas, L. O. and Rigler, D. **Influence of Friction in Making Settings on a Linear Scale** ... 434
Johnson, R. H. and Bond, G. L. **Reading Ease of Commonly Used Tests** ... 319
Jones, F. N. and Smith, C. J. **Visual Skill and Performance in a Meat Packing Plant** 313
Jones, M. H. **The Adequacy of Employee Selection Reports** ... 219
Jurgensen, C. E. **Intercorrelations in Merit Rating Traits** ... 240
Jurgensen, C. E. **Overall Job Success as a Basis for Employee Ratings** ... 333
Kates, S. L. **Rorschach Responses, Strong Blank Scales, and Job Satisfaction Among Policemen** ... 249
Keating, E., Paterson, D. G. and Stone, C. H. **Validity of Work Histories Obtained by Interview** ... 6
Kephart, N. C. and Besnard, G. G. **Visual Differentiation of Moving Objects** ... 50
Kephart, N. C. and Mason, J. M. **Acuity Differences between the Two Eyes and Job Performance** ... 423
Kerr, W. A. **Accident Proneness of Factory Departments** ... 167
Klugman, S. F. **Spread of Vocational Interests and General Adjustment Status** ... 108
Layton, W. L. **An IBM Card Profile for the Strong Vocational Interest Blank** ... 415
Levine, A. S. **Construction and Use of Verbal Analogy Items** ... 105
Levine, A. S. **Minnesota Psycho-Analogies Test** ... 300
Link, H. C. and Freiberg, A. D. **To What Extent Have the American People Accepted Socialism?** ... 88
Lord, F., Cowles, J. T. and Cynamon, M. **The Pre-Engineering Inventory as a Predictor of Success in Engineering Colleges** ... 30
Mandell, M. M. **The Administrative Judgment Test** ... 145
Miller, H. K., Jr. **An Exploratory Study of Linear Interpolation** ... 367
Mitchell, M. B. and Rothe, H. F. **Validity of an Emotional Key on a Short Industrial Personality Questionnaire** ... 329
McCormick, E. J. **An Analysis of Visual Requirements in Industry** ... 54
Nahm, H. **A Follow-Up Study on Satisfaction with Nursing** ... 343
Navran, L. **The Super-Roper Technique as a Measure of Interest in Nursing** ... 417
Nesberg, L. S. and Smith, K. U. **Measurement of a Complex Psychomotor Performance by Means of a Printed Test** ... 309
Owens, W. A. **An Aptitude Test for Veterinary Medicine** ... 295
Owens, W. A., Jr. **A Difficult New Test of Mechanical Comprehension** ... 77
Page, H. E. **A Note on Norms for the Purdue Industrial Mathematics Test and the Adaptability Test** ... 306
Pashalian, S. and Crissy, W. J. E. **How Readable Are Corporate Annual Reports?** ... 244

Paterson, D. G. **Report on the Journal of Applied Psychology for 1949** 285
Peatman, J. G. and Hallonquist, T. **Geographical Sampling in Testing the Appeal of Radio Broadcasts** 270
Poruben, A., Jr. **A Test Battery for Actuarial Clerks** 159
Pronko, N. H. and Herman, D. T. **Identification of Cola Beverages. IV. Postscript** 68
Prothro, E. T. and Perry, H. T. **Group Differences in Performance on the Meier Art Test** 96
Ramond, C. K., Rachal, L. H. and Marks, M. R. **Brand Discrimination among Cigarette Smokers** 282
Romney, A. K. **The Kuder Literary Scale as Related to Achievement in College English** 40
Ross, S. **A Study of Shooting Glasses by Means of Firing Accuracy** 118
Rothe, H. F. **Use of an Objectivity Key on a Short Industrial Personality Questionnaire** 98
Smith, A. J. **Menstruation and Industrial Efficiency. I. Absenteeism and Activity Level** 1
Smith, A. J. **Menstruation and Industrial Efficiency. II. Quality and Quantity of Production** 148
Stern, B. **Upper versus Lower Case Copy as a Factor in Typesetting Speed for Linotype Trainees** 351
Travers, R. M. W. and Wallace, W. L. **Inconsistency in the Predictive Value of a Battery of Tests** 237
Wagner, R. F. **Critical Requirements for Dentists** 190
Wallace, S. R., Jr. and Whitney, A. G. **The Prediction of Persistency in Premium Payment** 131
Weitz, J. **Verbal and Pictorial Questionnaires in Market Research** 363
Wesley, S. M., Corey, D. Q. and Stewart, B. M. **The Intra-Individual Relationship between Interest and Ability** 193
Wiener, D. N. and Simon, W. **Personality Characteristics of Embalmer Trainees** ... 391
Williams, A. C., Jr. and Roscoe, S. N. **Evaluation of Aircraft Instrument Displays for Use with the Omni-Directional Radio Range (VOR)** 123
Woods, W. A. and Boudreau, J. C. **Design Complexity as a Determiner of Visual Attention Among Artists and Non-Artists** 355

Book Reviews

Allport's The Individual and His Religion: Paul E. Meehl 451
Anastasi and Foley's Differential Psychology: Gladys C. Schwesinger 444
Bellows' Psychology of Personnel in Business and Industry: Albert S. Thompson 136
Bellows and Rush's Workbook in Personnel Methods: Albert S. Thompson 448
Bennett and Cruikshank's A Summary of Clerical Tests: Donald E. Super 214
Boynton's Selecting the New Employee: Robert N. McMurry 214
Brouwer's Student Personnel Services in General Education: Milton E. Hahn 72
Calhoon's Problems in Personnel Administration: Albert S. Thompson 444
Cavan, Burgess, Havighurst, and Goldhamer's Personal Adjustment in Old Age: Albert R. Chandler 217
Chapanis, Garner, and Morgan's Applied Experimental Psychology. Human Factors in Engineering Design: Jack W. Dunlap 213
Drake and Drake's A Human Relations Casebook for Executives and Supervisors: Albert S. Thompson 287
Flesch's The Art of Readable Writing: James J. Jenkins 71
Gilbert and Weitz's Psychology for the Profession of Nursing: Helen Nahm 139

Hoppock's Group Guidance; Principles, Techniques, and Evaluation: Milton E. Hahn 290
Hovland, Lumsdaine, and Sheffield's Experiments on Mass Communication. Studies in Social Psychology in World War II, Vol. III: Allen L. Edwards 139
Libo's Attitude Prediction in Labor Relations—A Test of Understanding: Edwin E. Ghiselli 371
Massing's Rehearsal for Destruction: Harrison G. Gough 447
Mathewson's Guidance Policy and Practice: Donald E. Super 141
McCord and Witheridge's Odors: Physiology and Control: Josef Brozek 447
Mossin's Selling Performance and Contentment in Relation to School Background: Daniel Raylesberg 287
Pease's Machine Computation of Elementary Statistics: Kenneth E. Clark 213
Prentice-Hall, Inc. The New Cure for White Collar Unrest: Clifford E. Jurgensen 138
Pressey's Educational Acceleration: Appraisals and Basic Problems: John W. Gustad 448
Reynolds and Shister's Job Horizons: C. E. Jurgensen 70
Schramm's Mass Communications: Alfred C. Welch 446
Stuit, Dickson, Jordan, and Schloerb's Predicting Success in Professional Schools: Barbara A. Kirk 445
Super's Appraising Vocational Fitness: J. R. Wittenborn 136
Tinkelman's Difficulty Prediction of Test Items: Charles I. Mosier 452
Van Der Lugt's V. D. L. Adult Psychomotor Test Series for the Measurement of Manual Ability and V. D. L. Psychomotor Test Series for Children for the Measurement of Manual Ability: Robert H. Seashore 371
Vernon and Parry's Personnel Selection in the British Forces: Dewey B. Stuit 288
Wagner's Selective Job Placement: A. A. Canfield 70
Ward's Out-of-School Vocational Guidance: Fred M. Fowler 450
Warner, Gardner, Henry, and Haggard's Identifying and Developing Potential Leaders: Ralph R. Canter, Jr. 289
Weitzman and McNamara's Constructing Classroom Examinations—A Guide for Teachers: Walter W. Cook 216
Weston's Sight, Light and Efficiency: Miles A. Tinker 449
Williamson's Trends in Student Personnel Work: Albert S. Thompson 72

Miscellaneous

New Books, Monographs, and Pamphlets 74, 143, 218, 292, 373, 454

Journal of Applied Psychology

VOL. 34, NO. 6 DECEMBER, 1950

The Concentric Organization Chart

C. G. Browne

Wayne University

The use of organization charts to illustrate personal and departmental relationships and the flow of authority and responsibility in any structured situation has long been a generally accepted procedure. Such charts perform the same functions with these non-quantified relationship variables as graphs do with quantified data. The organization chart also offers the same advantages of the graph, primarily the advantage of easier and quicker comprehension of the material under consideration.

It is the purpose of this article to present a type of organization chart diagramming, illustrated in Figure 2, which will be referred to as the concentric organization chart, and to indicate the advantages of this diagrammatic presentation of relationships over the more commonly used or traditional organization chart as illustrated in Figure 1. Both Figures 1 and 2 are based upon the organizational structure of a manufacturing company employing approximately 1,500 persons, all of the executives through the fourth echelon of the business being included in the charts illustrated.

It is proposed that, on the basis of the following considerations, the concentric organization chart is a more satisfactory representation of the material being presented than the traditional organization chart.

1. *Provides a better representation of the dynamics of personal relationships as they exist in an organizational structure.* The concentric organization chart represents dynamic, flowing relationships which proceed and are directed toward the center or inwardly, while at the same time there is a flow which proceeds and is directed from the center or outwardly. Characteristically, the functioning of any group or organization centers "around" individuals, not "below" them. This is clearly presented in the circular plan in Figure 2, culminating in this case with the President and General Manager as the focal point for the activities of the other executives involved. Or it might be said that individuals in a group, whether executive or supervisory or not, are surrounded with contacts, influences, and relationships that stimulate them from all directions. The schematic representation of these factors is better diagrammed as a circular rather than a horizontal plane. These flowing, surrounding, focalizing relationships then take on the desirable aspects of a smoothly functioning organizational unit in that they represent schematically the cooperation, intercoordination, and mutual dependency which characterize successful group activities.

2. *Eliminates the "above and below" concept.* The importance of the principles of general semantics is gaining recognition in psychology, particularly where it can be shown that human relationships and reactions are causally related to words as conceptual symbols. Many words, phrases, and concepts take on or are promoted for a certain emotional significance which influences the individual. Most individuals readily recognize various levels of authority and responsibility, and will accept their personal position in relation to these various levels. However, an emotional aspect attaches itself to this recognition when it becomes a reflection on the status of the individual in the concept of above or below, higher or lower, superior or inferior, up or down, top or bottom. The concentric organization chart eliminates this emotional concept, and diagrams organizational structure in such a way that each individual is a supporting part of the structure, the entire structure being dependent on all of these

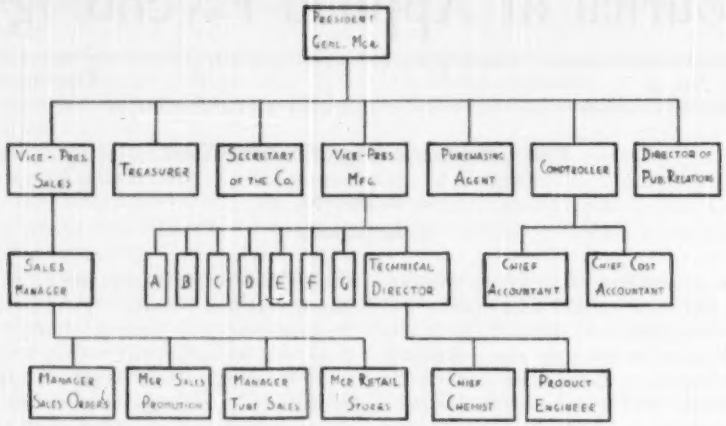


FIG. 1. Traditional type organization chart. Note: A. Mgr. Shipping; B. Mgr. Production Control; C. Mgr. Quality Control; D. Plant Engineer; E. Industrial Engineer; F. Personnel Director; and G. Foreman, Bicycle Tire Products.

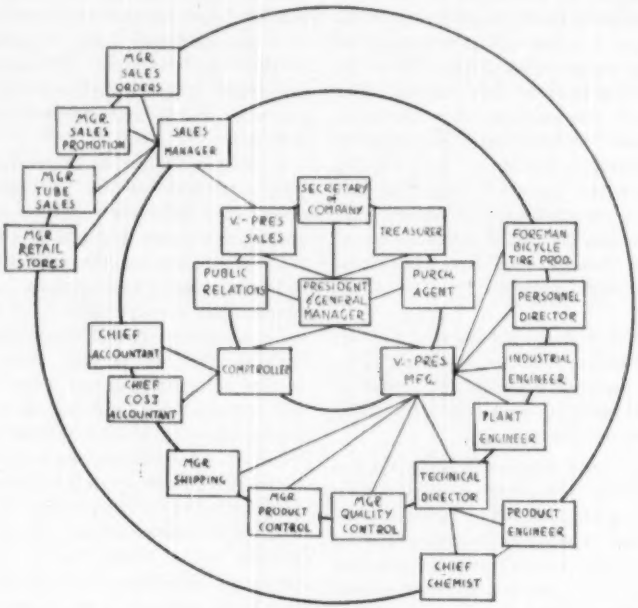


FIG. 2. Concentric organization chart. Note: Circles indicate echelon level; black connecting straight lines indicate flow of authority.

parts for its operation, since a break in any one of the concentric circles has an effect upon the entire organization and its functioning. Various echelons of authority are represented by the distance which any given circle is from the center or focal point.

3. *Presents an organization without loose ends.* An organization is a self-contained unit within the sphere of its own operations. However, the traditional organization chart does not represent this, unless some limits are superimposed on the chart itself. Instead, the ends or the right and left of each echelon level represent a drop-off or vacuum of the organization's activities which is not characteristic of the functioning of and the relationships existing within the organization. In effect, the concentric organization chart brings these loose ends around into circular form in a diagram which represents a self-contained unit functioning on a continuum within its own boundaries as contained in the chart, all of which portray the actual functioning of the organization.

4. *Eliminates the upside-down organizational structure.* Since the concentric organization chart has neither top nor bottom, the interpretations of the relationships existing within the organization are not dependent upon the position of the diagram which represents them. Any organization chart is designed to present certain facts about a given organization at a given time. These facts or relationships do not depend on viewing the organization from any particular angle or any specific position, which is also the case for the concentric organization chart. However, with the traditional organization chart, the relationships which it intends to portray can be interpreted properly only when the chart is presented to the viewer in a certain position, that is, with the top at the top and the bottom at the bottom. Should this not be done, the organization would be upside down, with relationships existing inversely to the facts of the organizational structure.

5. *Simplifies designing and understanding.* After the facts of the organization chart have been determined, there is a difficult problem of the arrangement of these facts on the diagram. This arrangement problem is particularly difficult on a traditional organization chart primarily because of the physical limitations of this type of chart. It is inevitable that the number of individuals will increase as the organizational set-up proceeds further and further from the focal point, or echelon one. That is, there is only one president, there may be four or five vice-presidents and other officers, 15 or 20 department heads, several hundred foremen, and thousands of non-supervisory personnel. However, on the traditional organization chart, the amount of space available for say the fifth or sixth echelon with 100 or more individuals is exactly the same space which is available for echelon one with one individual, the available space being a function of the width of the chart. This leads to many attempts and deviations by planners of organization charts to circumvent this problem in the best manner possible by changing the size of the box designating the individual (Figure 1, echelon 3), by raising or lowering the echelon level of some individuals in so far as the chart is concerned to an echelon where space is available (leading to confusion in following the relationships), by grouping individuals, or by any number of other devices.

The nature of the concentric organization chart automatically eliminates this primary problem of space in designing the chart. As illustrated in Figure 2, the necessity of including 11 executives on echelon three presents no problem since the third echelon circle is larger than the first and second echelon circles, where the number of executives is smaller. The same principle would follow in any expansion of the chart. This makes it possible to maintain a constant plan for showing the relationships between echelons and individuals, and for following these relationships through the system of connecting lines on the chart, which is not possible with the complicated net-work of lines and jumps to the left, right, up, or down which frequently create problems of interpretation and understanding with the traditional organization chart.

Received January 16, 1950.

Custom Made Systems of Job Evaluation

J. Stanley Gray

University of Georgia

Jobs have been evaluated since the beginning of employment but only in the last 25 years have any attempts been made to do this systematically. The term "job evaluation" now designates a *systematic*[1] attempt to weigh separately the worth of various elements or factors which constitute a job. This evaluation may be in money values (factor comparison), or point values (point systems), or some form of over-all classification. It is (or should be) preceded by job analysis, which also has been systematized in recent years. In other words, the trend in job evaluation and its concomitant activities has been toward systematization, with the assumption that more system will somehow result in greater validity.

This article presents one method of validating systems of job evaluation by elementary statistical procedure. Assuming that job evaluation is the determination of *relative* job values (i.e., the equitable distribution of a payroll—large or small) within an establishment, two other foundation theses are here presented. First, relative job values can be determined best by evaluating only those factors which are variables. If a factor exists in equal amount in all jobs being evaluated in an establishment, it is useless as a means of determining relative job values. This means that factors peculiar to an establishment must be used in evaluating jobs in that establishment. In other words *custom made* systems of job evaluation must be developed to fit each new job evaluation situation.

Second, the validity of any system of job evaluation (except factor comparison) can be determined by comparing its evaluation of key jobs with the wages already carried by those jobs. This means that when jobs are evaluated in points, the percentage of points as well as the percentage of wages for each key job can be calculated. It is then an easy matter to calculate the significance of the difference between the percentage of points and the percentage of wages carried by each of the key jobs.

On the basis of these assumptions the writer has developed and validated custom made systems of job evaluation for establishments in both steel and textile industries and for the non-academic jobs in all 16 institutions of the University System in Georgia. In each case, a job evaluation system was developed to fit the peculiarities of the situation. It was then tested on key jobs and modified until it evaluated these properly. In the industrial situations, two lists of key jobs were used to double-test the validity of the system.

The procedure used may be described in seven steps:

1. All jobs to be evaluated are first carefully analyzed and described. Any thorough schedule of job analysis is appropriate but the one developed by the War Manpower Commission and reported in the *Training and Reference Manual for Job Analysis*, U. S. Printing Office, 1944 is recommended. Too often job evaluators study only those factors included in their ready made system and ignore other important aspects of a job. When a job is analyzed before the evaluation factors are known, the analysis is likely to be more thorough and more appropriate for other uses, such as employment, training, etc.

2. The next step is a careful analysis of the job analyses to *discover* the important variables which are to become the evaluation factors. Some of the standard factors—working conditions, experience, hazards, physical effort—are investigated to see if they differentiate jobs from each other. If they are fairly uniform from job to job they are *not* used as evaluation factors. The number of factors that are used depends on the number of differentiating characteristics. Frequency tables are made for each factor investigated.

3. The differentiating factors thus discovered are then assigned basic weights according to the estimated importance of each. The writer arbitrarily uses 100 as the total

[1] The word "scientific" is often erroneously used instead.

basic weight for all factors. The value of each factor is then the per cent of its importance in relation to other factors. This constitutes the value of the first degree. Higher degrees are arithmetically increased by multiplying this base value by the degree number. Thus factor A (perhaps working conditions) may carry a weight of 10 as a base (first degree) and 20, 30, 40, etc. for higher degrees. This means that if the first degrees of all factors total 100, the second degrees will total 200, the third 300, etc. Most point systems of job evaluation determine factor and degree weights in this way. Illustrative weights are given in Table 1.

A part of the weighting step in developing a custom made system of job evaluation, and yet a part that should come only after the weighting has been completed, is the development of a manual. Each factor must be carefully defined, as well as each degree under each factor. A job evaluation manual is intended to be used as a guide so that the rating of jobs will not be difficult. How much of a trait a job may have should be easily decided by referring to the manual. Job evaluation manuals should be written in the most simple yet the most exact language possible. They should be brief yet not ambiguous. The various key jobs can be added later as typical examples of the degree meaning of each factor.

4. The fourth step is to select 15 or 20 key or bench mark jobs, that is, jobs which already carry proper pay rates. These may be selected in various ways, but it is well to have an advisory committee composed of individuals who are intimately acquainted with the jobs to be evaluated.[2] Men who have worked in an establishment for years are better able to select key jobs than is a job evaluator who has only a short and incomplete knowledge of job duties and even less acquaintance with pay rates. These *key jobs should be representative of all the jobs to be evaluated*. If some department has no job that is properly paid, in the opinion of the committee, a typical job can be selected and a fair wage assigned to it even though the actual wages are higher or lower. A key job with a hypothetical fair wage is just as good as one with an actual fair wage. With this system, jobs can be evaluated only as accurately as the rates on key jobs are accurate.

[2] Both management and labor should be represented on this committee. It can be used later to advise regarding the evaluation of more difficult and complex jobs.

Table 1

Factor and Degree Weights for a System of Job Evaluation

	Degrees				
	I	II	III	IV	V
1. Education	10	20	30	40	50
2. Experience	20	40	60	80	100
3. Physical Effort	10	20	30	40	50
4. Mental Effort	10	20	30	40	50
5. Working Conditions	10	20	30	40	50
6. Hazards	10	20	30	40	50
7. Responsibility for Safety	15	30	45	60	75
8. Tools, Materials, Equipment	15	30	45	60	75

5. Obviously the next step is to evaluate key jobs, using the table of factors and weights as developed in Step 3. When the total point value of each key job is determined, a table of evaluation points and wages should be constructed. See Table 2, columns 2 and 4. Then the per cent of total evaluation points as well as the per cent of total wages should be calculated for each key job. See Table 2, columns 3 and 5. The difference between the percentage of total evaluation points and the percentage of total wages carried by a job indicates the accuracy of the job evaluation. See Table 2, column 6. The statistical significance of this percentage difference should then be calculated. See Table 2, column 7. If this difference is significant, it is evidence that the job is *not* accurately evaluated and the table of factor weights must be adjusted. Only when all key jobs are evaluated in such a way that no difference between percentage of total evaluation points and percentage of total wages is statistically significant can the evaluation system be called valid.

6. It is well to double check a job evaluation system, especially if the point values of factors and degrees had to be juggled repeatedly to fit the key jobs. This necessitates a second list of key jobs chosen with the same care as the

Table 2

Key Jobs with Wage and Point Values

Jobs	Evaluation Points	Per Cent of Total	Hourly Wage	Per Cent of Total	Difference	S.D. of Difference	Difference / S.D. of Difference
Die Sinker A.	348	10.3	$1.95	10.5	0.2	.71	0.28
Hammersmith (Steam)	321	9.5	1.75	9.4	0.1	.68	0.15
Millwright (Maintenance)	305	9.0	1.65	8.9	0.1	.66	0.15
Machinist (Maintenance)	286	8.4	1.55	8.3	0.1	.64	0.16
Shaper Operator	275	8.1	1.50	8.1	0.0	—	—
Crane Hitcher	260	7.7	1.45	7.8	0.1	.62	0.16
Grinder (Cutters)	241	7.1	1.30	7.0	0.1	.59	0.17
Inspector (Bench)	229	6.8	1.25	6.7	0.1	.58	0.17
Scale Blower (Forge)	214	6.3	1.18	6.4	0.1	.57	0.17
Furnace Unloader	204	6.0	1.10	5.9	0.1	.55	0.18
Roll Mill Helper	190	5.6	1.05	5.7	0.1	.54	0.18
Tester (Brinell)	179	5.3	1.00	5.4	0.1	.53	0.19
Drill Press Opr.	172	5.1	0.95	5.1	0.0	—	—
Straightener	164	4.8	0.90	4.8	0.0	—	—
Total	3388	100.0	$18.58	100.0			

first. The jobs should be evaluated and a table constructed as described in Step 5. Again, if the differences between per cent of total points and per cent of total wages are great enough to be statistically significant, *the evaluation system is not yet valid.* Only when a job evaluation system justifies the wages of the key jobs can it be said to be valid enough to justify using it to evaluate other jobs which carry wage rates that may be too high, or too low, or just right.

7. The final step is to evaluate other jobs in the plant for which the system has been designed.[3] This means the preparation of an evaluation sheet for each job, a summary sheet showing the factor weights for all jobs, and a double entry graph showing the point values and present wage rates of all jobs. These results are then presented to the evaluation committee for their review and recommendations. These final steps are no different from those of any other job evaluation system.

[3] It must be remembered, however, that a custom made system of job evaluation is known to be appropriate for use with jobs typical of those represented in the key job sample. It may or may not fit jobs with different characteristics. If a group of jobs are found to be different from any one of the key jobs, it is then advisable to select the most typical one of them and add it to the other key jobs. This necessitates a repetition of Steps 4 and 5 and completely reworking the data tables. Often when jobs are greatly varied, it is necessary to develop more than one system of job evaluation. A system appropriate for evaluating production jobs is seldom appropriate for evaluating clerical jobs, or managerial jobs, or technical jobs. Different types of jobs are differentiated by different variables and should be evaluated by the use of appropriate factors.

Summary

Two innovations in job evaluation have been proposed in this paper. First, jobs are relatively evaluated by factors that represent differentiating characteristics between them. Ready made systems of job evaluation seldom fit because their factors do not differentiate between jobs. Valid job evaluation systems must fit the jobs being evaluated. Second, systems of job evaluation can be validated by computing the statistical significance of the differences between the evaluation of key jobs and the wage rates they carry. Key jobs are defined as those that already carry proper wage rates. When properly paid jobs are properly evaluated the evaluating system is then valid.

Received January 6, 1949.

Rating Training and Experience

Laverne K. Burke and Erwin K. Taylor

*Personnel Research Section, AGO**

One of the most expensive, laborious and time-consuming chores in the examining process of any merit system is the evaluation of training and experience of applicants. It is also one of the least reliable and least objective aspects of the entire examining process.

As the importance of the position under consideration increases there is generally a concomitant increase in both the weight given to background factors and in the difficulty of evaluation. The task frequently requires the services of subject-matter experts beyond the staff of most examining agencies. Where the number of applicants is large, difficulty in securing an adequate number of qualified raters with time to devote to the task frequently delays completion of the register for a matter of months. Additional problems are introduced by the need to train raters to achieve some degree of uniformity in standards and to assure the applicants of bias-free evaluations.

A task of similar nature faced the Department of the Army when, after the war, a need was felt to re-evaluate Army officers. To accomplish the job the Officers Efficiency Evaluation Board was appointed to place all officers in a relative order of merit. An Army Personnel Records Board made available to the Evaluation Board brief but accurate and complete data on each officer's commissioned service prior to 1 January 1947.

The Army Personnel Records Board developed a form which has become known as the Statement of Service. On it the Board recorded each assignment, rank, rating, promotion, transfer, decoration, and disciplinary action for the period from January, 1937 through December, 1946. Pertinent remarks of superiors were also extracted from the efficiency reports and recorded on the form. Data covering all service prior to 1937 were more briefly summarized.

Complete statements of service were sent to the Officers Efficiency Evaluation Board. This group was charged with the responsibility of reviewing the record and arriving at an evaluation of the individual officer's value to the Army. The task of reviewing and rating some twenty to thirty thousand statements of service was indeed formidable. The Personnel Research Section of the Adjutant General's Office was called upon for assistance and was charged with the development of an objective clerical procedure whereby an initial ordering of officers could be accomplished. Members of the board planned to examine this listing and the records and make any necessary adjustments required by factors not taken into account by the first approximation.[1] The Officers Efficiency Evaluation Board's final evaluation, called the War Service Score, correlated .86 with the system presented here.

Table 1

Estimated Reliability of Rankings, by Branch and by Sample

(In terms of average intercorrelation among 8 sets of 10 ranks each)

Branch	Sample A	Sample B
Cavalry	.80	.84
Coast Artillery	.90	.91
Engineers	.89	.83
Field Artillery	.90	.74
Finance	.68	.82
Infantry 1	.94	.93
Infantry 2	.88	.74
Ordnance	.88	.69
Quartermaster	.90	.83
Signal Corps	.83	.81

For purposes of analysis, a stratified sample of 200 lieutenant colonels was selected.

* The opinions expressed are those of the authors and do not necessarily represent official Dept. of the Army policy.

[1] The first approximation actually used was not the one here presented, but rather an improved system consisting of weighted composites of level of responsibility and efficiency reports.

Table 2

Correlation between Records Variables and Average Branch Rankings, Sample A

Note: All record variables, unless otherwise specified, are for the period from 1 Jan. 1937 through 31 Dec. 1946. Those marked by an asterisk are for the war period from 1 July 1943 through 30 June 1945.

Variable	Correlation Coefficient
1. Average efficiency rating	.81
*2. Average efficiency rating	.64
3. Average temporary grade held	.39
*4. Average temporary grade held	.56
5. Average: echelon of job X grade held	.39
*6. Average: echelon of job X grade held	.62
7. Per cent of war period spent overseas	.25
8. Age when Lt. Col. grade received	−.31
9. Present age	−.31
10. Number of times detailed to General Staff	.33
11. Total Army training received	.40
12. Highest Army training received	.44
13. Civilian education	.24
14. Years of enlisted service, not Regular Army	.09
15. Years of commissioned service, not Regular Army	−.09
16. Years of enlisted service, Regular Army	.02
17. Years of commissioned service, Regular Army	.02
18. Hospitalized, neurosis	.01
19. Hospitalized, physical	−.18
20. Highest grade reached	.69
21. Average temporary grade held (1942–1946)	.68
22. Highest efficiency rating	.15
23. Lowest efficiency rating	.69
24. Range of efficiency rating	−.68
25. Highest echelon assignment	.41
26. Total decorations received	.69
27. Highest decoration received	.67
28. Total number of assignments (1937–1941)	−.19
29. Total number of assignments (1942–1946)	−.13
30. Number of months unassigned, non-hospital	−.20
31. Number of months unassigned, hospital	−.13
32. Demotions	.63
33. Mention of special aptitudes by superiors	.23
34. Height	.01
35. Weight	−.06
36. Weight-height ratio	−.11

Twenty were selected at random from each of eight branches and 40 from infantry. The sample from each branch was sub-divided into two groups, so that one stratified sample of 100 could be used for developing the evaluation system and the other comparable sample could be used for checking or cross validation.

For each branch subsample of ten cases the statements of service and special ranking forms were sent to the respective Career Management Branches. In each of these branches the ten officers in each subsample were *independently* ranked by eight officers who studied the statements of service carefully. To determine the reliability of these independent rankings, the average intercorrelation among rankers by branch was computed. These are presented in Table 1.

Careful analysis of the statement of service revealed 36 variables capable of quantification.

The remarks made by superiors when rating the officers were examined for the number and type of descriptive words and phrases used. However, consideration of the tendency of some superiors to put more descriptive remarks in their rating reports than others, even when rating the same officer, plus the length of time required to tally and check the number of different words used indicated that the analysis of the narrative material would not yield variables adequately reliable to justify their inclusion in the study.

The intercorrelations among the 36 variables from the statements of service and the average branch rankings were computed for 98 officers[2] in sample A. Table 2 presents the correlation of each of the 36 variables with the average rank. Table 3 presents the intercorrelations among the 36 variables.[3] Applying the Wherry-Doolittle Test-Selection technique,[4] a battery consisting of variables: (1) average efficiency report; (20) highest grade reached; (23) lowest efficiency report; and (27) highest decoration received, was developed.

The correlation of this battery with average

[2] Two officers in this sample had been prisoners of war and evaluation of their war service could not be made.

[3] To reduce printing costs and to save space Table 3 has been deposited with ADI and may be ordered as Document 2749 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $0.50 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $0.50 for photocopies (6 × 8 inches) readable without optical aid.

[4] Stead, Shartle, & Associates. *Occupational Counselling Techniques.* New York: Am. Book Co., 1940. Pp. 245–252.

Table 4

Correlation between Branch Rankings and Composite of 4 Records Variables, by Branch and by Sample

Note: The combining weights were derived from the data of sample A and were used in combining the four variables for both samples.

Branch	Sample A	Sample B
Total	.90	.85
Cavalry	.93	.93
Coast Artillery	.90	.98
Engineers	.97	.92
Field Artillery	.98	.78
Finance	.77	.82
Infantry 1	.99	.82
Infantry 2	.94	.98
Ordnance	.93	.82
Quartermaster	.97	.93
Signal Corps	.79	.98

rank by branch is given in Table 4. To make sure that the high correlations in sample A (on which the test selection was based) were not merely a function of sampling error, the battery was also applied to the independent sample B. The correlation of battery scores with average ranking for sample B is also presented in Table 4.

It is interesting to note that there is better agreement between the battery score and the average ranking than there is between pairs of competent evaluators working from the same basic data. This fact argues strongly for the use of a mechanical system of evaluating background items. Such a system not only results in considerable savings of time, money, and personnel but results in an evaluation that is consistently more reliable than the more laborious method of subjective evaluation.

Summary

As a possible means of evaluating Army officers' merit the Personnel Research Section developed a simple mechanical system of evaluation whereby four variables are objectively rated by reference to a set of specified weights. The system was cross-checked against a second group of officers and found to be effective. The evaluations made by the system agreed more closely with the consensus of groups of eight competent evaluators than pairs of evaluators agreed with each other in the subjective rating of the officers' entire records.

It is suggested that the technique employed could be of considerable service to merit systems. We estimate that it would be worthwhile doing when there are in excess of a thousand applicants to be rated. The system has the advantage that it will usually be possible to secure much more competent evaluators to make the trial rankings than could be secured to rate the entire applicant population. It is thus thought that application of the method described will not only save time and effort but will result in better evaluations as well.

Received February 1, 1950.

Measuring the Level of Abstraction

Rudolf Flesch

Dobbs Ferry, N. Y.

Students of communication agree that awareness of the level of abstraction is essential for full comprehension. For example, Perrin, in his college text, says (**14**): "For exact and reasonable communication it is highly important that a speaker or writer knows where in the range of meaning of abstract words his core of meaning falls and that he makes this clear to his listeners or readers." Semanticists are especially emphatic on this point. Johnson (**10**) writes: "The prime objective of general semantics is to make one conscious of abstracting." Hayakawa (**8**) writes: "Consciousness of abstracting is . . . a sign of adulthood."

Though the significance of the level of abstraction is widely recognized, no studies have been reported that attempt to estimate or measure it quantitatively. Discussions are usually limited to generalized distinctions between abstract and concrete words, illustrative examples, and such figurative devices as the semanticists' "abstraction ladder."

The present study is an attempt to approach this problem with the technique developed for the measurement of readability.

In recent years a number of statistical formulas to estimate readability (comprehension difficulty) have been developed (e.g., **3**, **12**), among them a four-part formula reported by the writer in 1948 (**4**, also **5**). That formula was based on the following style elements: (1) average sentence length in words; (2) average word length in syllables; (3) average percentage of "personal words"; and (4) average percentage of "personal sentences."

To adapt the technique of readability measurement to the measurement of the level of abstraction, the writer utilized the fact that certain parts of speech are more frequent in abstract expression, while certain other parts of speech are more frequent in concrete expression. It is well known, for instance, that an abstract style contains relatively more descriptive adjectives, indefinite pronouns, and subordinating conjunctions, while a concrete style contains relatively more proper nouns, limiting adjectives, finite verbs, personal pronouns, and coordinating conjunctions. Most of these statistical facts have never been used for the measurement of readability, but some of them have been the basis of such related studies as Boder's development of the adjective-verb quotient (**1**), Fries' analysis of the language of the Bible (**7**), LaBrant's study of subordination in children's writing (**11**) and Séchéhaye's study on the logical structure of the sentence (**18**). The present study is an attmept to combine and extend these statistical analyses of parts of speech and bring them in line with currently accepted grammatical classifications, as exemplified in the works of Curme (**2**) and Jespersen (**9**).

To sum up the hypothesis underlying the present study: Level of abstraction can be estimated by computing the ratio of certain parts of speech to certain other parts of speech in written expression. Since level of abstraction is a basic element in readability (comprehension difficulty), this ratio can be used as a measure of readability, either by itself or in combination with other elements.

Procedure

The criterion used in the writer's earlier readability formula (**4**) was McCall-Crabbs' *Standard test lessons in reading* (**13**). Since this criterion is keyed to children's grade levels, it seemed also a useful criterion for measuring the level of abstraction. It is generally agreed that children's mental growth corresponds to a growing ability to handle abstractions (see, for instance, Piaget, **15**, **16**, **17**).

To make the prediction as accurate as possible, the following 42 of the 376 McCall-Crabbs' passages (containing poetry, directions, or arithmetical problems) were omitted from the count: Book II, Nos. 4, 42, 45, 77, 89, 92, 93, 94; Book III, Nos. 6, 11, 14, 15, 22, 32, 36, 41, 46, 51, 52, 63, 76, 78, 82, 84, 92, 94; Book IV, Nos. 8, 17, 31, 37, 43, 51, 75; Book V, Nos. 1, 21, 32, 52, 64, 79, 80, 83, 94.

To test the level of abstraction, the writer analyzed the generally accepted grammatical

categories of the parts of speech. It was found that most parts of speech contained certain categories that were statistically related to abstractness and certain others related to concreteness. In general, words related to abstractness are more indefinite, those related to concreteness more definite. The writer, therefore, chose the arbitrary label "definite words" for the words whose percentage was used to measure concreteness (level of abstraction).

The categories of "definite words," arrived at by a process of trial and error, were these (for more detailed definitions, see the section "How to Use the Formula" below):

(1) The following nouns: Common and proper nouns with natural gender; common and proper nouns specifying time; nouns in the possessive case ending in 's or s'; and nouns modified by the limiting adjectives listed below.

(2) The following limiting adjectives: Possessive adjectives; intensifying adjectives; numeral adjectives; the adjectives *what, this, these, that, those, each, same, both;* and *the* if modifying a noun not otherwise modified.

(3) Finite verbs, except the verb *to be* used as copula.

(4) Present participles ending in *ing* if used to form the progressive tense.

(5) The following pronouns: Personal pronouns; reflexive pronouns; the relative pronouns *who, whose, whom, what, that;* interrogative pronouns; and the listed limiting adjectives when used as pronouns.

(6) The following adverbs: *here, there, now, then, where, when, why, how.*

(7) All interjections.

(8) The words *yes* and *no.*

The percentage of these "definite words" in the test passages was correlated with the grade level of children who answered correctly one-half of the test questions.

To combine the new count of "definite words" with other tests of readability, intercorrelations were then computed. This yielded a multiple regression formula to predict readability, based on the percentage of "definite words" and the average word length in syllables. Following the procedure in the earlier study (**5**), this combined readability formula was then expressed in such a way that maximum readability had a value of 100 and minimum readability a value of 0.

Findings

The correlation between the percentage of "definite words" and the average grade level of children who could answer correctly one-half of the test questions was found to be $r = -.5541$. The corresponding regression formula is: $C_{50} = 8.7493 - .0852\ dw$.

On the basis of this evidence, the percentage of "definite words" may be considered useful to form a rough estimate of the level of abstraction, ranging from 0 (fully abstract) to 100 (fully concrete).

The percentage of "definite words" may also be combined with average word length in syllables to form a measure of readability. The intercorrelations, means, standard deviations, and regressions weights found are shown in Tables 1 and 2. The following symbols were used: *wl* for word length (syllables per 100 words), *dw* for percentage of "definite words," C_{50} for the average grade of children who could answer one-half of the test questions correctly, and C_{75} for the average grade of children who could answer three-quarters of the test questions correctly.

The regression formula based on these correlations is: $R = 168.095 + .532dw - .811\ wl$ (R represents the readability score).

Scores computed by this formula have a range from 0 to 100 for almost all samples

Table 1

Correlations, Means, Standard Deviations, and Regression Weights of Word Length and "Definite Words"

	dw	C_{50}	$\bar{X}$	s	β
wl	−.5681	.6908	135.1437	14.0030	.5552
dw		−.5541	37.9790	9.1720	−.2387

Table 2

Means and Standard Deviations of Two Criteria

	$\bar{X}$	s
C_{50}	5.5135	1.4096
C_{75}	7.1302	2.0454

taken from ordinary prose. A score of 100 corresponds to the prediction that a child who has completed fourth grade will be able to answer correctly three-quarters of the test questions to be asked about the passage that is being rated; in other words, a score of 100 indicates reading matter that is understandable for persons who have completed fourth grade and are, in the language of the U. S. Census, barely "functionally literate." The range of 100 points was arrived at by multiplying the grade level prediction by 10, so that a point on the formula scale corresponds to one-tenth of a grade. However, this relationship holds true only up to about seventh grade; beyond that, the formula underrates grade level to an increasing degree. Finally, the formula—which predicted grade level and, therefore, difficulty—was "turned around" by reversing the signs to predict readability. (Before this transformation the formula read: $C_{75} = .0811\ wl - .0532\ dw - 1.8095$.) The multiple correlation coefficient of this formula is $R = .72$.

Since the correlations with the criterion C_{50} were higher than those with the criterion C_{75}, the multiple correlation with the criterion C_{50} was computed first; as a second step, the value so found was used to predict criterion C_{75}, in order to predict 75% comprehension rather than 50% comprehension.

Comment

The percentage of "definite words" appears to be a useful test in two ways:

First, it is a rough measure of the level of abstraction. As such, it may be used as a tool in semantic studies, critical reading, literary appreciation, translation, rating of advertising copy, and propaganda analysis. Step-by-step analysis of the level of abstraction in a given piece of discourse may also be helpful in logical analysis and the discovery of faults in reasoning.

Second, the new test is a measure of readability. As such it replaces the two tests of "human interest" (percentage of "personal words" and "personal sentences") that formed part B of the writer's earlier formula (**5**). Combined with the average word length in syllables, it gives a practical and comprehensive measure of readability.

It should be noted that this measure approaches the problem of readability in a new way, since it emphasizes the importance of the newspaperman's "Five W's" (Who? What? Where? When? Why?), the role of names and addresses, facts and figures, and the use of practical examples, illustrations, and anecdotes. Negatively, it points up the bad effects of too many adjectives, abstract nouns, and weak verb forms. It may, therefore, be expected to become a useful training device in the teaching of composition.

According to Shannon's mathematical theory of communication (**19**, **20**), any noise hampering the transmission of a message is counteracted by redundancy in the message itself. By extension, any "semantic noise" (comprehension difficulty, ambiguity, etc.) is counteracted by "semantic redundancy" (repetition, padding, amplification, restatement, etc.). In other words, information on a high level of abstraction, which is difficult to grasp, is communicated more effectively to the extent that restatements on a more concrete level are added. This, of course, corresponds to the use of illustrations, practical applications, examples, parables, and the like in teaching. The new measure of abstraction may possibly be used in conjunction with the mathematical theory of communication to estimate the number and level of concrete illustrations necessary to convey a given piece of abstract information to a given audience.

This general principle has long been recognized in practice and there is a marked tendency in modern writing to alternate between abstract generalizations and concrete examples. Because of that, the percentage of "definite words" often fluctuates between extremes within a single piece of writing. This may make more or longer samples necessary in order to arrive at a reliable estimate.

The correlation of the new factor to comprehension (.55) is higher than that of average sentence length to comprehension (.52), as reported in the writer's earlier paper (**5**). In other words, concreteness appears to be somewhat more important for readability than the use of short sentences.

An amusing illustration of the new measure is furnished by Hayakawa (**8**), who explains his "abstraction ladder" by way of four statements on different levels of abstraction. Ap-

plication of the new measure to these examples shows the following:

	Ratio of "definite words"	Percentage of "definite words"
Mrs. Plotz makes good potato pancakes.	3 in 6	50
Mrs. Plotz is a good cook.	2 in 6	33
Chicago *women* are good cooks.	1 in 5	20
The culinary art *has* reached a high state in America.	1 in 10	10

How to Use the Formula

For practical application, the directions for using the formula may be stated as follows:

To measure the level of abstraction of a given piece of writing, go through the following steps:

Step 1. Unless you want to test a whole piece of writing, take samples. Take enough samples to make a fair test. Don't try to pick "good" or "typical" samples. Go by a strictly numerical scheme. For instance, take every third paragraph or every other page. Each sample should start at the beginning of a paragraph.

Step 2. Count the words in your piece of writing or, if you are using samples, take each sample and count each word in it up to 100. Count contractions and hyphenated words as one word. Count as words numbers or letters separated by space.

Step 3. Count the following "definite words." Count each of these words only once. Count as words all units separated by white space. (In the examples, "definite words" are italicized.)

(1) Count all names of people—that is, proper nouns with natural gender, either masculine or feminine. Count all words that are part of the name, including titles, etc., used as part of the name. Example: *President Harry S. Truman* (count 4).

Count names of people used as adjectives to modify natural-gender nouns (e.g., the *Smith brothers*, the *Dolly sisters*), but do not count personal names used as adjectives to modify nouns *without* natural gender (e.g., the Ford Motor Company, the Washington Monument).

(2) Count all common nouns that have natural gender, either masculine or feminine. Examples: *father, mother, iceman, actress.* Count these words also when used as adjectives to indicate the gender of another noun (e.g., *woman doctor, bull elephant, girl athlete*), but do not count them when they do not indicate gender (e.g., fellow workers).

Do not count common-gender nouns like teacher, doctor, employee, spouse.

(3) Count all nouns that indicate a specific time on the clock or the calendar; e.g., the names of the months, the seasons, the days of the week, the words *morning, noon, afternoon, evening, night, day* (in the sense of daytime), *sunup, sundown, today, tonight, yesterday, tomorrow*, and the words *breakfast, lunch, dinner, supper* when used to indicate time. Count these words also when used as adjectives to specify time (e.g., *December day, fall season, lunch hour*), but do not count them when used as adjectives without relation to time (e.g., Thursday club, dinner companion, Sunday suit).

(4) Count all numeral adjectives and all nouns modified by numeral adjectives. Count also the words *first, next, last*, and such words as *double, pair, half, triple.* Examples: *27 words. 54 percent. Six letters.* The *next day.* The *last moment.* But don't count nouns that are not directly modified by numerals, e.g., *thousands* of people, a *ten-year-old* house.

Count the word *one* in the sense of "single," but not when used as an indefinite pronoun or as part of "no one," "any one," or "some one." Examples: Count *one* in "*one* fine morning" and in "the pretty *one*," but not in "one has one's doubts."

Count the word *once* in the sense of "a single time," but not in the phrase "at once." Examples: *Once* upon a time. *He succeeded once.*

Count the words *other* and *another* only in the sense of "second" or "one more."

(5) Count all finite verb forms—that is, verbs in the first, second, or third person and the present, past, or future tense. In verb forms with auxiliary verbs, count the auxiliary rather than the main verb. Examples: *He came* and *went. We have* considered. *You should* have declined. *Let us* pack up and go. *It was* debated and voted down.

Exception: Do not count the verb "to be" when used as copula, that is, as a link between subject and predicate. Do not count "to be" as copula in any form, with or without an auxiliary verb. Examples: *I* was sick. *You* should have been *there. It* might be fun. *He* ought to have been careful.

(6) Count all present participles (*-ing*) when used as part of the progressive tense (to be *-ing*). Examples: *He was running. I am going* to look. *You should* have been *working.*

(7) Count all personal pronouns except indefinite "it," and all reflexive pronouns, formed with *-self* or *-selves.* Examples: It is a fact that *you* and *I* are not related. It was *Mary herself.*

(8) Count the words *here, there, then*, and

now (except indefinite "there" in "there is," "there are," etc.).

(9) Count the words *who, whom, when, where, why,* and *how.*

(10) Count the words *this, that, these,* and *those; each, same,* and *both;* and nouns modified by them.

(11) Count the words *what* and *which* (interrogative) and nouns modified by them, but not the word "which" when used as a relative pronoun. Examples: *Which way are you going?* But: The car which *I bought.*

(12) Count all possessive pronouns (*my, your, his, her, its, our, their,* etc.), all nouns in the possessive case ending in *'s* or *s'*, the word *whose,* and all nouns modified by these possessives. Examples: *Our* modern *civilization, its* recent *development, whose business, journey's end.*

But do not count possessive cases of pronouns not otherwise counted, e.g., "one's ideas," "someone else's hat."

(13) Count the word *that* when used as a relative pronoun, but not when used as a conjunction. Examples: Remember that *you* are sick. It was the humidity *that did* it.

(14) Count the words *yes* and *no* (used as answer).

(15) Count all interjections.

(16) Count the definite article *the* and the noun modified by it, but only if that noun is a single word not otherwise modified. Examples: *I missed the bus. We called the doctor. It* was *the truth.* But: *I missed* the green bus. *We called* the eye specialist. *It* was the truth, plain and simple. *They gave me* the room *I had last time. It* was the beginning of *the end. We live* at the end of *the village,* in the yellow house at the turn of *the road.*

Do not count the word "the" when modifying a noun that is to be counted under one of the other definitions. Examples: *He came* in the *afternoon.* The only *one there* was the *boy.*

Do not count the word "the" when modifying adjectives or noun-adjectives, particularly proper noun-adjectives referring to nationality, race, etc. Examples: *Their team* was the best. *We sat* in the dark. *What's* the good of *it?* The Scotch are thrifty. The Negro *was* arrested.

Step 4. The number of "definite words" in your 100-word sample (or the average percentage of "definite words" in all your samples or the whole piece of writing tested) indicates the level of abstraction. The typical relation between the percentage of "definite words" and level of abstraction is shown in Table 3.

To test the readability of a piece of writing, go through the following additional steps:

Step 5. Count the syllables in your 100-word samples or, if you are testing a whole piece of writing, compute the number of syllables per 100 words. If in doubt about syllabication rules, use any good dictionary. Count the number of syllables in symbols and figures according to the way they are normally read aloud, e.g., two for $ ("dollars") and four for 1918 ("nineteen-eighteen"). If a passage contains several or lengthy figures, your estimate will be more accurate if you don't include these figures in your syllable count. In a 100-word sample, be sure to add instead a corresponding number of words in your syllable count. To save time, count all syllables except the first in all words of more than one syllable and add the total to the number of words tested. It is also helpful to "read silently aloud" while counting.

Table 3

Typical Percentages of "Definite Words"

Level of Abstraction	Percentage of "Definite Words"
Highly abstract	Up to 20
Fairly abstract	20 to 30
Fairly concrete	30 to 45
Highly concrete	Over 45

Step 6. Find your readability score by inserting the number of "definite words" per hundred words (*dw*) and the number of syllables per hundred words (word length, *wl*) in the following formula:

$$R = 168.095 + .532\ dw - .811\ wl.$$

The readability score will put your piece of writing on a scale between 0 (practically unreadable) and 100 (easy for any literate person). Readability scores will tend to follow the pattern shown in Table 4.

Table 4

Pattern of Readability Scores

Readability Score	Description of Style	Syllables per 100 Words	"Definite Words" per 100 Words
0 to 30	Very difficult	192	5
30 to 50	Difficult	167	14
50 to 60	Fairly difficult	155	22
60 to 70	Standard	147	29
70 to 80	Fairly easy	139	36
80 to 90	Easy	131	43
90 to 100	Very easy	123	50

Or you may want to test a piece of writing separately for concreteness and "reading ease." If so, do this after *Step 5:*

Alternative Step 6: Figure the average sentence length in words for your piece of writing or, if you are using samples, for all your samples combined. In a 100-word sample, find the sentence that ends nearest to the 100-word mark—that might be at the 94th word or the 109th word. Count the sentences up to that point and divide the number of words in those sentences by the number of sentences. In counting sentences, follow the units of thought rather than the punctuation: usually sentences are marked off by periods; but sometimes they are marked off by colons or semicolons—like these. But don't break up sentences that are joined by conjunctions like *and* or *but*.

Alternative Step 7: Find your "reading ease" score by inserting the number of syllables per 100 words (word length, *wl*) and the average sentence length (*sl*) in the following formula:

R. E. ("reading ease")
$$= 206.835 - .846\,wl - 1.015\,sl.$$

The "reading ease" score will put your piece of writing on a scale between 0 (practically unreadable) and 100 (easy for any literate person).

Sample Application

To show the application of the new test, three passages dealing with the subject of truth will be used. Passage A, exemplifying a highly concrete style, is Test Lesson No. 37 from Book III of the *Standard Test Lessons in Reading* (**13**) that were used as criterion. Passage B was taken from *Courts on Trial* by Jerome Frank (**6**). It is an example of writing on the middle level of abstraction, illustrating the more abstract generalizations in the first paragraph by more concrete examples in the second. Passage C was taken from *The Next Development in Man* by L. L. Whyte (**21**) and shows a highly abstract, philosophical discussion of truth.

"Definite words" in the three passages are italicized.

Passage A:

The children were telling about *their Christmas vacations.*

"*We went* to Kansas," *said Jack.* "*One day when we were skating* on *the lake* some of the *boys cut* a hole in *the ice, struck* a match and a fire *blazed* right up out of *the hole* for *two* or *three minutes.*"

"*Oh, oh!*" *said* all the others, "*that* couldn't be true. Water *doesn't* burn."

"But *it* is true," *said Jack.* "*I saw it.*"

They turned to *the teacher* to see *what she would* say and *she explained this* very strange *happening.* It *seems* there are natural gas wells under *the lake* which *send the gas* bubbling up through *the water where it is* caught in large pockets under *the ice.*

"So *you see,*" *said the teacher,* "*when* a hole *is* cut the escaping gas *will* burn if lighted."

Passage B:

No means *then, have* as yet been discovered, or are likely to be discovered, for ascertaining whether or to *what extent* the belief of the trial judge about the facts of a case *corresponds* to the objective facts as *they* actually *occurred, when the witnesses disagree,* and *when* some of the oral testimony, taken as true, *will* support the *judge's conclusion.* In other words, in such a case there is no objective measure of the accuracy of a *judge's finding* of *the facts.* There *exists* no yardstick for *that* purpose.

In a "contested" law suit, therefore, with *the witnesses* in disagreement, usually no one *can* adequately criticize the trial *judge's fact-finding.* If, at the end of *the trial,* the trial judge *says* that *Jones hit Smith,* or that *Mrs. Moriarity called Mrs. Flannagan* a liar, or that old *widow Robinson* was insane *when she made her will,* or that *Wriggle used* fraud in inducing *Simple* to sign a contract—the *judge's word goes.* And the *same* would be true if, in most of *those instances,* the trial judge *had* found exactly the opposite to be *the facts.*

Table 5

Analysis of Three Sample Passages

	Percentage of "Definite Words"	Syllables per 100 Words	Readability Score
Passage A	53 (highly concrete)	121	98 (very easy)
Passage B:			
First paragraph	24 (fairly abstract)	151	58 (fairly difficult)
Second paragraph	37 (fairly concrete)	152	65 (standard)
Both paragraphs	31 (fairly concrete)	151	62 (standard)
Passage C	17 (highly abstract)	159	48 (difficult)

Passage C:

Truth is thought which *conforms* to the form of the whole. Conformity to the whole is *the criterion*. The unitary truth is *that* which *conforms* to the whole process of which *it* is a part. *The truth* is a form embedded in the whole complex of processes in the human organism and *its environment*, symbolizing and organizing *them*. A particular truth *may* not represent the entire structure of a situation but only *those aspects* which are relevant to thought at a given stage in *its development*. *The truth* is a system of symbols *whose structure conforms* to the whole pattern of feeling, thought, and action, and *integrates* all the processes which *link* the reception of stimuli and the molding of the ultimate responses. *This* is not a pragmatic criterion, since unitary truth *does* not merely serve special needs but *unites* the whole system in a conviction which is at once emotional, intellectual, and practical.

A comparative analysis of the three passages is given in Table 5.

Considerable experience with practical application of the new test has shown that an untrained person can acquire reasonable familiarity with this test after ten to twenty applications and will then be able to count the "definite words" in a 100-word sample in one to two minutes. The application of the rest of the readability test will ordinarily take another one to two minutes.

Received September 26, 1950.
Early publication.

References

1. Boder, D. P. The adjective-verb quotient: a contribution to the psychology of language. *Psychol. Record*, 1940, **22**, 310–343.
2. Curme, George O. *Principles and practice of English grammar*. New York: Barnes & Noble, 1947.
3. Dale, Edgar, and Chall, Jeanne S. A formula for predicting readability. *Educ. Res. Bull.*, Ohio State Univ., 1948, **27**, 11–20, 28.
4. Flesch, Rudolf. *The art of readable writing*. New York: Harper & Brothers, 1949.
5. Flesch, Rudolf. A new readability yardstick. *J. appl. Psychol.*, 1948, **32**, 221–233.
6. Frank, Jerome. *Courts on trial; myth and reality in American justice*. Princeton Univ. Pr., 1949.
7. Fries, Charles C. One stylistic feature of the 1611 English Bible. *Fred Newton Scott Anniversary papers*. Univ. of Chicago Pr., 1929, 175–187.
8. Hayakawa, S. I. *Language in thought and action*. New York: Harcourt, Brace & Co., 1949.
9. Jespersen, Otto. *Essentials of English grammar*. New York: Henry Holt & Co., 1933.
10. Johnson, Wendell. *People in quandaries; the semantics of personal adjustment*. New York: Harper & Brothers, 1946.
11. LaBrant, Lou L. A study of certain language developments of children in grades 4 to 12 inclusive. *Genet. Psychol. Monogr.*, 1933, **5**, 387–491.
12. Lorge, I. The Lorge and Flesch readability formulae: a correction. *Sch. & Soc.*, 1948, **67**, 141–142.
13. McCall, W. A., and Crabbs, Lelah M. *Standard test lessons in reading*. Books II, III, IV, and V. New York: Bur. of Publ., Teachers Coll., Columbia Univ., 1926.
14. Perrin, Porter G. *Writer's guide and index to English*. Chicago: Scott, Foresman & Co., 1942.
15. Piaget, Jean. *The child's conception of causality*. New York: Harcourt, Brace & Co., 1930.
16. Piaget, Jean. *Judgment and reasoning in the child*. New York: Harcourt, Brace & Co., 1928.
17. Piaget, Jean. *The language and thought of the child*. New York: Harcourt, Brace & Co., 1926.
18. Séchéhaye, Albert. *Essai sur la structure logique de le phrase*. Paris: Champion, 1926.
19. Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.*, 1948, **27**, 379–423, 623–656.
20. Weaver, Warren. The mathematics of communication. *Scientific American*, July 1949, **181**, No. 1, 11–15.
21. Whyte, Lancelot L. *The next development in man*. New York: Henry Holt & Co., 1948.

Personality Characteristics of Embalmer Trainees

Daniel N. Wiener and Werner Simon, M.D.

Veterans Administration, Fort Snelling, Minnesota *

Occasionally clients are counseled who have a strong desire to be embalmers or undertakers. This vocational objective is seldom suggested to a client unless he shows high original motivation to enter this field.

In a regional veterans' counseling program thirty-six veterans indicated their desire to become embalmers and began vocational training. The vocation is recognized through standards set by the State of Minnesota, and the organized training program includes two years of college work prior to specialized training.

Psychiatrists and psychologists, in discussing the occupation of embalming, frequently raise the question as to personality traits of those attracted to this field. In an attempt to investigate this question a study has been made of the personality characteristics of veterans who indicated their interest in this work, in order to determine whether they are obviously unique in personality. This is but one of many occupations where the technical skill and social attitudes involved may be less important factors to the counselee than his personal peculiarities in motivation and needs.

Literature

In a review of the recent literature of psychiatry and psychology for possible prior research in this area, no study of any kind has been found, although there are a number of analyses on concepts of death, as well as several case reports on necrophilia. The literature on death concepts has only secondary importance here. Perhaps it is sufficient to say that the fear of death, which appears so commonly in reports of dreams and conscious anxieties, is seldom discussed openly. This apparently repressed concern is indirectly reflected, however, in a number of ways; e.g., public responsiveness to "fear" campaigns to raise funds for medical research (cancer, infantile paralysis, tuberculosis), and in the elaborate rituals of funerals. In relation to the present problem this literature on concepts of death suggests an investigation of the personality structure of those who seem to lack this common fear as shown by their work with the dead (**2**, **3**).

Case reports on necrophilia stress pathological personality organization, especially in the sexual sphere. Brill (**1**) reports the case of a thirty-year-old single male who was primarily a passive homosexual and whose necrophilia represented a by-product of his pathological sexuality. He gave a history of a severe phobia with regard to corpses since early childhood, which was apparently overcome by touching a corpse. This experience changed the former phobia into the opposite, a strong attraction for corpses. As his fascination for corpses increased, he became the paid assistant of an undertaker, who spoke highly of his diligence and skill. During analysis he confessed that it was his morbid desire to play with corpses which prompted him to work for the undertaker.

Rapoport's (**4**) patient was a fifty-year-old man who was arrested because he was found kissing and touching a female corpse in a funeral parlor. He began his practice of visiting dead bodies six years after the death of his mother, and a ritual of visiting funeral parlors two or three times a week ensued. The corpses were usually female, middle-aged or elderly. Rapoport points out that the literature contains very little on the subject of necrophilia, but it seems to him unlikely that the number of cases is as small as the absence of literature suggests.

Method

It seemed desirable to determine common impressions of psychologists and psychiatrists as to what personality peculiarities, if any, embalmers might have. This was done by asking both vocational and personal counselors, as well as psychiatrists, for their impressions, and tabulating their responses.

* Published with permission of the Dean's Committee, Medical School, University of Minnesota.

The actual personality characteristics of embalmer trainees were measured by use of the Minnesota Multiphasic Personality Inventory, with the recently developed subtle and obvious keys (5) included. A group of thirty-six veterans were studied, all of whom indicated a desire to be embalmers and made this their training objective after counseling. The control group consisted of one hundred veterans who came to the same guidance centers for counseling during the same period of time, but who decided upon other vocational objectives. All were tested with the MMPI.

The use of the term "embalmer" may be distinguished from that of "undertaker." The undertaker receives training first in embalming, followed by additional training involving more of the business and social details of funeral arrangements.

Results

Thirteen vocational and personal counselors were asked to write briefly their impressions of the personality characteristics of men interested in embalming. Thirty-six responses were elicited, with a range of one to five reactions from each counselor. Table 1 indicates the results of the tabulation of the responses. Relatively shallow emotionality, as well as aggressiveness, are the outstanding impressions, with feminine characteristics also suggested.

Table 1

Personality Characteristics of Embalmers According to Thirteen Personal and Vocational Counselors

Traits	Frequency
Psychopathic	11
High Drive	9
Feminine	5
None, Stable	5
Introvertive	3
Scientific Interest	2
Depresaive	1
	36

Five psychiatrists gave ten responses to a similar question, with a range of one to three items. On tabulation, responses classified as "extrovertive," "compulsive," "passive-aggres-

Table 2

Means and Standard Deviations of Embalmer (N = 36) and Control (N = 100) Groups in Age and Education, and on MMPI

		Mean	S.D.
Age	Emb.	25.72	4.66
	Con.	24.48	4.24
Education	Emb.	12.00	1.55
	Con.	11.68	1.63
K	Emb.	56.45	7.85
	Con.	55.15	9.60
F	Emb.	50.20	4.90
	Con.	50.95	5.80
Hypochondriasis*	Emb.	54.25	8.55
	Con.	51.00	9.55
Depression**	Emb.	49.65	10.05
	Con.	52.35	11.40
D-Subtle	Emb.	48.70	11.75
D-Obvious	Emb.	51.30	9.20
Hysteria	Emb.	55.90	10.60
	Con.	55.35	8.55
Hy-Subtle	Emb.	57.70	9.50
Hy-Obvious	Emb.	52.05	10.50
Psychopathic Dev.	Emb.	55.50	8.40
	Con.	55.54	11.80
Pd-Subtle	Emb.	55.50	8.95
Pd-Obvious	Emb.	50.35	8.15
Femininity**	Emb.	54.80	8.70
	Con.	56.90	7.45
Paranoia*	Emb.	47.85	8.50
	Con.	50.75	8.55
Pa-Subtle	Emb.	51.75	10.90
Pa-Obvious	Emb.	44.00	6.15
Psychasthenia	Emb.	55.10	6.75
	Con.	54.44	10.10
Schizophrenia	Emb.	56.35	8.00
	Con.	56.14	9.45
Hypomania**	Emb.	59.55	11.90
	Con.	56.80	11.45
Ma-Subtle	Emb.	61.30	11.70
Ma-Obvious	Emb.	51.75	9.80

* Difference between means is significant at 2–3% level.

** Difference between means is significant at 6% level.

sive" and "none" were given twice. There were single responses for "depressive" and "psychopathic."

Table 2 gives the means and standard deviations for the embalmer trainees and control groups, together with an indication of the significance of the difference between the means of each MMPI Scale.

In age and education the embalmer and control groups were not significantly different from each other. In Hypochondriasis, Depression, Femininity, Paranoia and Hypomania, however, the means do differ slightly from each other with some likelihood of significance. Most prominent is the slightly greater tendency toward Hypochondriasis in the embalmers. Hypomania is somewhat less elevated, while in Depression, Femininity and Paranoia the embalmers are slightly lower than the control group. Most of the subtle scores are higher than the obvious scores, a characteristic of successful groups in previous studies (**6**).

Summary

These test results indicate that there are no strongly deviate personality characteristics among the embalmer trainees. Elevations on the subtle scores, along with the Hypomania score, seem to characterize successful groups in other areas (**6**). The fact that the scores for several other scales are somewhat lower for embalmers than for the control group suggests the same conclusion.

The strongest possibility for an interpretation of uniqueness lies in the elevated Hypochondriasis. Tendencies toward over-concern with personal health may, through projection, lead to an interest in the dead and in preparing bodies for burial.

Negatively, the results of this study do not substantiate tendencies toward shallow emotionality, femininity, compulsiveness or necrophilia, which a strong interest in dead bodies might suggest and which both counselors and psychiatrists tend to indicate.

Received January 23, 1950.

References

1. Brill, A. A. Necrophilia. *J. crim. Psychopath.*, 1941, **2**, 433–443.
2. Caprio, F. S. A psycho-social study of primitive conceptions of death. *J. crim. Psychopath.*, 1943, **5**, 303–317.
3. Orlansky, H. Reactions to the death of Pres. Roosevelt. *J. soc. Psychol.*, 1947, **26**, 235–266.
4. Rapoport, J. A case of necrophilia. *J. crim. Psychopath.*, 1942, **4**, 277–289.
5. Wiener, D. N. Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1948, **12**, 164–170.
6. Wiener, D. N. A control factor in social adjustment. *J. abnorm. soc. Psychol.* (In press.)

The Guilford-Zimmerman Temperament Survey and Certain Related Personality Tests *

Claudia Gilbert

The Pennsylvania State College

As a result of an accumulation of evidence obtained by testing supervisors in industrial plants in Pennsylvania, the Personnel Service Division of The Pennsylvania State College is now using the Bernreuter Personality Iventory and the Guilford-Martin Personnel Inventory in test batteries for the selection of candidates for supervisory positions. Consequently, the publication of a new personality test, the Guilford-Zimmerman Temperament Survey, which appears to incorporate in one test the traits now being measured by a combination of the Guilford-Martin and the Bernreuter, is of great interest to the Division.

The purpose of the present investigation was to set up an experimental situation which allowed a comparison of the Guilford-Zimmerman, Guilford-Martin, and Bernreuter, and an analysis of the resulting intercorrelations, in order that the Personnel Service Division might have a starting point for further experimentation with this new test.

The Tests

Since the purpose of this study was to make a practical comparison of the Bernreuter with the Guilford-Zimmerman, it was decided to use the three most widely applied Bernreuter scales—B1-N-Neurotic Tendency (reversed and oriented as Stability), B2-S-Self-Sufficiency, and B4-D-Dominance—these also being the scales used by the Personnel Service Division in selection for supervisory trainees.

The Guilford series of personality tests are the result of an extensive series of factor analyses that have identified a number of personality factors. Of the original Guilford series of three personality scales, the Guilford-Martin Personnel Inventory measuring Objectivity, Agreeableness, and Cooperativeness has been found to offer good possibilities in industrial situations. The authors have published studies which give an indication of its validity, particularly in regard to the selection of supervisors (**5**).

The new Guilford-Zimmerman Temperament Survey measures ten Guilford traits termed GRASE-OFTPM. These ten traits, measured with 300 items by the Guilford-Zimmerman, have been arrived at by means of condensations and omissions of trait scores where intercorrelations are sufficiently high, thus purporting coverage in a more economical fashion of the 13 factors measured by the three original Guilford inventories. Rather than being composed of questions, the new inventory is made up of declarative statements, so that instead of "Do you get things done in a hurry?" it is, "You get things done in a hurry." No validation studies of the Guilford-Zimmerman Temperament Survey are known to the author.

Procedure

The subjects for the study were the members of two introductory psychology classes at The Pennsylvania State College. Of these, an all male population of 64 was used. Scatter diagrams were prepared in order to ascertain the linearity of the data. The Pearson product moment formula was employed and, in all, 120 correlations were determined.

Results

Intercorrelations of the Traits. In Table 1 are presented the coefficients of correlation, in terms of product moment *r*'s, between the Guilford-Zimmerman factors and their counterparts in the Guilford-Martin. O-Objectivity in the Guilford-Zimmerman correlated .78 with O-Objectivity in the Guilford-Martin; F-Friendliness in the Guilford-Zimmerman correlated .76 with Ag-Agreeableness in the Guilford-Martin, and P-Personal Relations in

* This material is derived from the author's thesis submitted in June, 1950, in partial satisfaction of the requirements for the M.S. degree at The Pennsylvania State College. The author is deeply indebted to Charles Griffin and Dr. Kinsley R. Smith for their guidance and criticism.

Table 1

Coefficients of Correlation *

		Guilford-Zimmerman				
		O-Objectivity	F-Friendliness	P-Personal Relations	E-Emotional Stability	A-Ascendance
Guilford-Martin	O-Objectivity	.78				
	Ag-Agreeableness		.76			
	Co-Cooperativeness			.82		
Bernreuter	B1-N (rev.) Stability				.80	
	B4-D-Dominance					.80

* All *r*'s significant at the .01 level.

the Guilford-Zimmerman correlated .82 with Co-Cooperativeness in the Guilford-Martin.

Also in Table 1 are presented the coefficients between the Guilford-Zimmerman factors, E-Emotional Stability and A-Ascendance, and traits B1-N (reversed) Stability and B4-D-Dominance in the Bernreuter. E-Emotional Stability in the Guilford-Zimmerman correlated .80 with B1-N (reversed) Stability in the Bernreuter, and A-Ascendance in the Guilford-Zimmerman correlated .80 with B4-D-Dominance in the Bernreuter.

Further Analysis of Intercorrelations. In extending the comparison between the factors in the Guilford-Martin and their counterparts in the Guilford-Zimmerman to include comparison with the remaining seven factors in the Guilford-Zimmerman as well as with the Bernreuter traits, several correlations fell into the category of *r*'s generally accepted as denoting high to very high relationship (4). A correlation of .74 was found between O-Objectivity in the Guilford-Martin and B1-N (reversed) Stability in the Bernreuter. Also, there was a correlation of .74 between this same factor O-Objectivity in the Guilford-Martin and E-Emotional Stability in the Guilford-Zimmerman.

In comparing the Bernreuter traits with the Guilford factors, the only correlation between these measures which could be accepted as high, aside from those already discussed, was the .74 coefficient between B1-N (reversed) Stability in the Bernreuter and O-Objectivity in the Guilford-Martin.

The intercorrelations found in this study when comparing the Guilford-Zimmerman traits to each other, in terms of product moment *r*'s, indicated similar interrelationships to those indicated by the tetrachoric coefficients published in the Guilford-Zimmerman manual. There was one very high coefficient, however, which fell well above our arbitrary standard for a high correlation. This was a .84 between E-Emotional Stability and O-Objectivity.

The high intercorrelation between B1-N (reversed) Stability and O-Objectivity in both the Guilford-Martin and the Guilford-Zimmerman brings to mind the high intercorrelation of −.95, as reported in the Bernreuter Manual, between B3-I-Introversion and B1-N-Neurotic Tendency. This suggests, as does the data presented in this paper in terms of correlation coefficients, that on the one hand B1-N, oriented as Stability and B3-I, oriented as Extraversion, measure the same thing, and on the other hand, that this is the same trait as the Guilford-Zimmerman and the Guilford-Martin factor O-Objectivity. The high correlation of .84 between O-Objectivity and E-Emotional Stability in the Guilford-Zimmerman again suggests that these are not separate traits.

To sum up, the evidence presented suggests that the various personality inventories which are measuring traits named Objectivity (or Extraversion) and Emotional Stability, are primarily measuring only one factor.

A Comparison of B2-S-Self-Sufficiency with the Guilford Factors. There was no indication of a relationship between B2-S-Self-Sufficiency and any of the Guilford factors. The highest correlation was, in fact, .37, which though significant at the .01 level of confidence, appeared to be more noteworthy for its lowness than for its highness. However, such a result seems understandable, since up to this point the trait Self-Sufficiency has not been named by Cattell, Guilford, or Thurstone, and apparently has not been isolated in factor analysis. Also, it remains the most indepen-

dent measure, in terms of intercorrelations, in the Bernreuter Personality Inventory.

On the other hand, there could be another explanation. An inspection of the scatter diagrams of B2-S with the Guilford-Zimmerman factors suggests that in regard to several factors, especially A-Ascendance, there may be curvilinearity rather than linearity of relationship. In terms of the present research, this means the relationship between the Guilford factors and B2-S may have been masked when studied in terms of the Pearson *r*, and significantly higher correlations might have appeared if the *eta* correlation were applied. To check this assumption the *eta* between B2-S-Self-Sufficiency and A-Ascendance was computed. The chi-square test for linearity of regression indicated, however, that the departure of curvilinear relationship from linear relationship was not significant at the .05 level.

Summary

A study was made of the relationships between the factors measured by the Guilford-Zimmerman Temperament Survey and the traits measured by two previously validated inventories, the Guilford-Martin Personnel Inventory and the Bernreuter Personality Inventory. The subjects for the study were members of two introductory psychology classes at The Pennsylvania State College. Of these, an all male population of 64 was used. Scatter diagrams were prepared in order to ascertain the linearity of the data. The Pearson product moment formula was employed and, in all, 120 correlations were determined.

1. O-Objectivity in the Guilford-Zimmerman has a high positive correlation ($r = .78$) with O-Objectivity in the Guilford-Martin.

2. P-Personal Relations in the Guilford-Zimmerman has a high positive correlation ($r = .82$) with Co-Cooperativeness in the Guilford-Martin.

3. F-Friendliness in the Guilford-Zimmerman has a high positive correlation ($r = .76$) with Ag-Agreeableness in the Guilford-Martin.

4. E-Emotional Stability in the Guilford-Zimmerman has a high positive correlation ($r = .80$) with B1-N-Neurotic Tendency reversed and oriented as Stability, in the Bernreuter.

5. A-Ascendance in the Guilford-Zimmerman has a high positive correlation ($r = .80$) with B4-D-Dominance in the Bernreuter.

6. There is no indication of a high degree of relationship between B2-S-Self-Sufficiency on the Bernreuter and any of the Guilford factors.

7. The high positive correlation between the Guilford-Zimmerman factors O-Objectivity and E-Emotional Stability of .84 suggests that these are not separate traits. There is further evidence for this in the high positive correlation between O-Objectivity in the Guilford-Martin and E-Emotional Stability in the Guilford-Zimmerman of .74, and between O-Objectivity in the Guilford-Martin and B1-N-Neurotic Tendency reversed and oriented as Stability in the Bernreuter of .74.

Received June 26, 1950.
Early publication.

References

1. Allport, G. W. Neurotic personality and traits of self-expression. *J. soc. Psychol.*, 1930, **1**, 524.
2. Bernreuter, R. G. *Manual for the Personality Inventory.* Stanford University, California: Stanford University Press, 1935.
3. Cattell, R. B. Confirmation and clarification of the primary personality factors. *Psychometrika*, 1947, **12**, 197–220.
4. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans, Green and Company, 1948.
5. Guilford, J. P. *Manual for the Guilford-Martin Personnel Inventory.* Beverly Hills, California: Sheridan Supply Company, 1943.
6. Guilford, J. P., and Zimmerman, W. S. *Manual for the Guilford-Zimmerman Temperament Survey.* Beverly Hills, California: Sheridan Supply Company, 1949.
7. Mosier, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika*, 1937, **2**, 268–287.
8. Super, D. E. The Bernreuter Personality Inventory: a review of research. *Psychol. Bull.*, 1942, **39**, 94–125.
9. Thurstone, L. L. The factorial description of temperament. *Science*, 1950, **111**, 454–455.

The Non-Respondent Problem in Questionnaire Research

W. Leslie Barnette, Jr.

Department of Psychology, University of Buffalo

A great deal of social science research is necessarily conducted by means of mail questionnaires, despite wide recognition of pitfalls with such methodology. It is imperative, especially when questionnaire returns obtained on a sample are to be utilized for predictions for a population—the usual goal—that information be available about the non-respondents.

No attempt will here be made to review the literature which disucsses this topic. The interested reader is referred to three recent summaries (**5**, **6**, **12**). Most workers attempt to minimize non-respondent bias by maximizing their returns, usually through repeated follow-ups. Toops' early reports (**9**, **10**) are well known in this connection. Here he was interested in securing a 100 per cent response, something which may rarely be attempted when the original sample is very large. Others (**3**, **7**, **8**) have tried to cope with the problem by using special appeals and even bribes. Despite the fact that these techniques are effective, one is still left with a sizeable "hard core" non-respondent group. It becomes imperative to assess the possible source of bias here.

It is not always the case that the more verbal individuals respond. Clark's study (**4**), involving a mail questionnaire to A. F. L. union members, demonstrated this. The experience of the present writer has been similar (**1**), during a pre-testing of a questionnaire to be used on a large group of veterans. Here some 275 questionnaires were mailed out to a random sample; three weeks later a 50 per cent response was secured with no intervening follow-up. Just as many replies were received from veterans in professional and managerial categories as from those in skilled and semi-skilled groups.

In the opinion of the writer, the only *real* solution is by taking a random sample of the entire non-respondent group. By repeated and persistent follow-ups one aims to secure, ideally, a 100 per cent return. One then has relevant data for the detection of non-respondent bias. The following experience of the writer is offered as evidence.

A large-scale follow-up of some 1,300 veterans in the Greater New York area was in progress. Replies to the first questionnaire mailing totalled 580; a second mailing, some three months later, produced 310 additional replies. The combined first two waves thus gave 890 (69 per cent) respondents. The average time lapse between the date last seen and the questionnaire receipt was two years. All addressees had received free educational and vocational counseling at the agency in question so that they were presumably in its debt. First versus second waves of returns, when checked by chi-square, showed no important differences. However, despite this comforting fact, there still remained a total of 409 (31 per cent) non-respondents.

It was not practical to attempt a third contact of the entire 409. A random sample of the group was drawn (every fifth name from an alphabetical file, which gave $N = 82$). It was not then a matter of merely sending out a third letter and waiting for the subsequent wave of returns. Persistent attempts were necessary with this group by means of intervening postcard reminders, further letters, telephone calls where feasible and, in some cases, personal calls at homes.

Parenthetically, the unit cost of such a procedure is prohibitive. For many of this group, a total of five separate personal letters were involved, each time inclosing the questionnaire and a stamped, self-addressed envelope; for others, suburban New York telephone calls, two and three per person. Such could rarely be done for as large a sample as the entire 409 recalcitrants. The net result was an expenditure of a great amount of time and effort and money for, in the main, a relatively small amount of data in return.

The spontaneous comments of some of these individuals showed that the repeated reminders by mail, the tone of which became more sharp each time, were effective. One respondent,

apparently ashamed by the persistent reminders, finally returned the completed questionnaire, after receipt of the fifth letter, together with five three-cent stamps, as if in payment for the previously unsuccessful efforts at contact.[1] Another respondent, after receipt of the sixth letter, returned the completed questionnaire with an apologetic note: ". . . thanks for having such patience with me and for making me ashamed of myself to a degree where I find myself resolving not to procrastinate in any form or manner whatsoever." Few, however, were as courteous or gave any signs of such remodelled behavior.

In the end, returns were received from all but three persons in this random sample (i.e., 96 per cent replies). The results of this non-respondent check need not be reported here. Suffice it to say that, in the main, the trends observed in the data from the first two waves were corroborated. A few important differences were evidenced, however, which permitted a more reliable estimate for the entire veteran sample than would otherwise have been possible. These results are presented in two forthcoming papers by the present writer (2).

A subsidiary check for *respondent* bias (i.e., "early" versus "late" returns), which might be a cue for the presence of non-respondent bias, is occasionally available to the researcher. It is never one that might be employed on a large scale, however, without involving considerable annoyance to addressees. When one is engaged in sending out a second and third letter to non-respondents, it sometimes happens that one draws two returns from the same individual. One may then compare these two sets—a type of test-retest technique. The present writer met this situation for only five respondents; there was a minimum of three months' time lapse between the receipt of the first and second questionnaire. The average number of items on the questionnaire responded to by these persons was 12. The mean per cent of exact duplication in answers was 92. The instances of discrepancies between replies from the first versus second questionnaire all related to salary data, all of which involved a salary raise for the intervening time.

[1] The respondent was in error by 50 per cent; he should have returned thirty cents.

Summary

It is suggested that the real solution to the problem of bias detection, in the event that a numerically large non-respondent group remains after two or three mail questionnaire follow-ups, is to make up a random sample of this hard core; then to make repeated and persistent attempts at contact. The aim here is a 100 per cent response. An illustration of the application of this technique has been given in connection with a large-scale follow-up of 1,300 veterans where, after two mailings, there still remained some 400 non-respondents. An incidental check, of the test-retest type, is pointed out.

Received February 28, 1950.

References

1. Barnette, W. L., Jr. Preliminary report on a follow-up of veterans counseled at the Vocational Service Center, Y. M. C. A. of the City of New York. *Counseling* (Y. M. C. A., 347 Madison Avenue, New York City 17), 1948, **6** (no. 2), 1–2.
2. Barnette, W. L., Jr. Report of a follow-up of counseled veterans. I. Public Law 346 versus Public Law 16 clients. II. Status of pursuit of training. *J. soc. Psychol.* (forthcoming).
3. Bevis, J. C. Economical incentive used for mail questionnaire. *Publ. Opin. Quart.*, 1948, **12**, 492–493.
4. Clark, K. E. A vocational interest test at the skilled trades level. *J. appl. Psychol.*, 1949, **33**, 291–303.
5. Clausen, J. A., and Ford, R. N. Controlling bias in mail questionnaires. *J. Amer. stat. Ass.*, 1947, **42**, 497–511.
6. Norman, R. H. A review of some problems related to the mail questionnaire technique. *Educ. psychol. Measmt.*, 1948, **2**, 235–248.
7. Rollins, M. The practical use of repeated questionnaire waves. *J. appl. Psychol.*, 1940, **24**, 770–772.
8. Sletto, R. Pretesting of questionnaires. *Amer. sociol. Rev.*, 1940, **5**, 193–200.
9. Toops, H. A. Validating the questionnaire method. *J. Person. Res.*, 1923, **2**, 152–169.
10. Toops, H. A. The returns from follow-up letters to questionnaires. *J. appl. Psychol.*, 1926, **10**, 92–101.
11. Toops, H. A. Questionnaires. Pp. 948–951. In Monroe, W. S. (ed.): *Encyclopedia of Educational Research*. New York: Macmillan, 1950 (revised).

Reactions of Veterans to Counseling *

W. Leslie Barnette, Jr.

Department of Psychology, University of Buffalo

As part of a large-scale follow-up of counseled veterans by means of a mail questionnaire, the reactions of such clients to the counseling process were solicited. These data provide considerable information regarding the value of counseling as seen from the vantage point of the client and have diverse implications for the psychological counselor. At the bottom of a one-page questionnaire, designed primarily to elicit facts about the veteran's current status in training some two years after last seen, two items were placed which directly touched on client attitudes to services received. A pilot study, in which the questionnaire had been pre-tested, produced many letters from veterans amplifying the information requested on the questionnaire itself and dealing specifically with counseling reactions. These proved to be of such value that, when it came to the main study, efforts were made to encourage as many of these spontaneous comments as possible.

Two mailings of the questionnaire were involved. The working total (after deduction of post office rejects) was 1,299. From the two mailings, a total of 890 respondents was secured, containing approximately twice as many non-disabled veterans (those covered by the provisions of Public Law 346) as disabled (Public Law 16). To check on the reliability of the obtained data from these two follow-ups, a random sample of the non-respondents was selected for persistent and repeated follow-up. A 96 per cent return from this non-respondent sample was secured, the results of which altered the main trends of the data in no important way. It is therefore felt that the data from the study of the 890 respondents from the first two waves possess considerable reliability.

The respondents were all completed advisements from the files of the Vocational Service Center of the Y. M. C. A. of New York City (hereafter, VSC). The two questions dealing with attitudes towards counseling were adapted from the reports of the Adjustment Service (**6, 10**). The first of these asked: Did you consider the suggestions made to you at VSC to be (check one) helpful, of doubtful value, impractical? The second: I found the attitude of the VSC staff to be (check one) sympathetic, indifferent, discourteous. The distribution of the replies from the 890 returns is given in Table 1.

Table 1

Veterans' Opinions of VSC

Suggestions were:		
Helpful	668	(75%)
Doubtful	169	(19%)
Impractical	28	(3%)
No reply	25	(3%)
Totals	890	100
Staff attitude was:		
Sympathetic	809	(91%)
Indifferent	56	(6%)
Discourteous	0	
No reply	25	(3%)
Totals	890	100

Despite the highly favorable opinion, replies falling into the other categories were welcomed as evidence that the returns were not producing a kind of "white-wash." Many respondents appended further comments in support of this majority opinion, two samples of which are:

> Glad I visited your office. Put me on the goal to my future. If not for your counselors I guess I never would enjoy working today. Thanks for your kind service.
>
> . . . Personally I found your counseling so impressive that I resolved to "stick to it" regardless of future distractions, inducements, etc., believing that would be the best, most rewarding, long-range plan.

A total of 19 per cent indicated the counselor suggestions were of doubtful value. These

* This article is based on a section of the author's Ph.D. dissertation in the Department of Psychology, New York University, entitled "Occupational Aptitude Patterns of Counseled Veterans." All statements relating to veterans are the sole responsibility of the author. Colonel Walter Ketcham, Chief of Advisement of the New York Regional Office of the Veterans Administration, approved the entire research project.

persons, in the main, either regarded the suggestions as unrealistic because the training suggested would be too extensive or else they indicated disappointment at being given no "new" information about themselves.

A still smaller group (3 per cent) checked the "impractical" category. In support of these ratings one states:

. . . Suggestions impractical because the suggested goal requires years of night study of an extremely complex subject (law) without any possibility of an adequate living for the years of study plus other years of attempting to establish a clientele; and I was and am unwilling to begin life at 40 . . . (continues about his present job). . . . Am doing work I enjoy, am making a living and have time to worry about the world and its works. In other words you can put me down as "adjusted" even though your recommendation was not followed. Thanks for your help.

Turning to a brief consideration of client reactions to the item concerning the staff attitude, the overwhelming majority is again highly favorable. Many of the unsolicited comments were overly enthusiastic, containing direct compliments and testimonials to counselor effectiveness. The following quotations, from the many, will give the flavor.

I thank you and the staff from the bottom of my heart. May there be more people like you. Keep up the good work.

Was most favorably impressed by the helpful attitude and results. At no time was the impression given that here was a group of magicians who could determine an exact endeavor which could not fail—yet the pattern made by the apparently exhaustive tests was very clear. Can't suggest anything more helpful than the hope you'll continue—as is.

A very small percentage indicated they felt some indifference. This is rarely explained and, when attempted, is somewhat garbled. No respondent checked the "discourteous" entry—possibly too strong a word here.

More productive from the point of view of studying counseling from the vantage point of the client are the spontaneous remarks in connection with these two items.[1] This type of supplementary data, of course, was not received from all. Of the total of 890 returned questionnaires, 575 individuals or 64 per cent offered additional remarks. These comments run a large gamut. They vary in length from a succinct thanks for help to a 13-page letter, from notes full of errors in grammar and spelling to typed biographical essays, from a warm and receptive attitude to the questionnaire and the whole idea of a follow-up to irate negativistic attitudes,[2] from well-adjusted and secure individuals to persons with signs of aggravated emotional instability. All these comments were not limited to counseling matters; it will be only these types, however, that are discussed in this paper. Most of the commenting group supplied amplification concerning their job or educational history; others gave extensive occupational information; others raised questions that demanded answers; still others simply talked about irrelevancies. Any such classification of these diverse comments must be a subjective one. The categories which proved useful are reported in Table 2. Any one respondent may have made more than one

Table 2

Content Analysis of All Spontaneous Remarks (N = 575)

Info. augmenting Q items (i.e., no new facts)	392 (70%)
Incr. self-confidence	80 (14%)
Specific VSC compliments	100 (17%)
Thanks for your interest	152 (26%)
VSC criticisms	102 (18%)
VSC recommendations made	25 (4%)
Preoccupied with tests	89 (15%)
School difficulties	64 (11%)
Job difficulties	65 (11%)
Employer gripe	16 (3%)
Feels goal unrealistic	29 (4%)
Raises questions for answer	64 (11%)
VA criticisms	12 (2%)
Miscellaneous	69 (12%)

[1] The covering letter which accompanied the questionnaire specifically requested any type of additional information, such as feelings about VSC, that the veteran would like to supply. The letter also encouraged frankness and used the altruistic appeal that such information would be of great value in helping VSC deal with current veteran clients.

[2] The most extreme example of this is from a Public Law 16 client with a 50 per cent disability for psychoneurosis, anxiety. He returned a blank questionnaire with the following comments: "What are you guys doing over there? Pimping for the VA? Last time I had anything to do with you, all I got was a lot of red tape and a 20 per cent cut in my pension. So from now on my name is KIZ MIAZ."

type of comment; the percentages will not therefore total 100.

Increased Self-Confidence

Beyond a general statement from many respondents indicating an appreciation for the interest and help extended, many mention the fact that tests and counseling served to increase self-confidence. To vocational counselors, this is no great surprise. Seipp's report (**10**) for the Adjustment Service noted the same result; Stone and Simos (**11**), in their Minnesota follow-up, likewise. Despite the fact that no specific question was asked regarding this point, 80 (14 per cent) of individuals making spontaneous comments specifically mention this point.

> Tests encouraged me to continue with prewar program; confidence increased. . . . Saved the trouble of finding out the hard way.
>
> . . . In helping me realize what action to take, you and your staff opened a bright road in front of me, for the road I had traversed before was always dark and had no hope to see a light at the end of it.

Critical Comments on Counseling

Respondents were not uniformly so commendatory. On the questionnaire some 22 per cent indicated they regarded the suggestions given as of doubtful value or impractical. Some of these went on to a fuller explanation of such attitudes. While many such comments evidence great misapprehension as to the functions of a counseling agency, at least most are honestly offered and, in the main, well meant. These remarks frequently betray an exaggerated feeling of dependence towards psychological help. These respondents feel VSC is not definite enough in its suggestions, while others think it is too definite; others think too much emphasis is placed on interests, claimed or measured. And others are annoyed that the tests and counseling did not provide them with startlingly novel information about themselves. Some feel not enough pressure is brought by the counselor—to "force decisions," as one respondent phrased the matter.

> . . . Tests and interviews left me just about where they found me. Confirmed obvious aptitudes; showed up my inaptitudes. Got no actual guidance. When left the building, was just as befuddled as to the next step as when started. No schooling or type of job was recommended. . . . Perhaps in helping a party such as me, it is harder to point out the road to follow. . . . In my case perhaps I had to find the machinery or equipment I *wanted* to sell. . . . Only an experienced person in touch with that type of business could have helped me. . . .
>
> . . . Staff courteous but not convincing enough. Should actually force the individual to try to decide whether he likes an occupation or not. . . . Why not arrange for actually visiting different occupations by potential employees? . . . What good is testing a man whose test reveals he is best suited for a closed or already over-crowded field: spread out—don't limit yourself to one small part of the job.

Recommendations Offered

A few respondents—25 in all or some 4 per cent of the comment group—tried to make specific suggestions as to how the service might be improved. Some of these suggestions come from people who had distinctly favorable attitudes toward VSC but who felt the job might be done still better; other suggestions came from respondents with definitely unfavorable points of view. Actually there were only a small number of constructive suggestions.

The two most frequent recommendations are that the counselors be more specific, do more "telling" and that a placement service be installed. The former is in disagreement with much of the best counseling theory; the latter is now an accomplished fact at VSC. Other suggestions are occasionally sound but there are others that are very unrealistic—those which advocate even more authoritarian approaches to clients. In the former category are suggestions to the effect that more time should be devoted to actual counseling—more discussion of the individual, more "analysis" done. These are the persons who, while they regard the tests as valuable, want to go considerably beyond them. Pressure of time and the case load per counselor frequently operate against this. Certainly, however, wherever this aspect could be lengthened it would be desirable; it would also serve to reduce the ever-present potential of counseling failures.

Test Preoccupied Clients

There has been previous mention of signs of psychological dependency in some respondents,

especially those with negatively critical comments about services rendered. There is another—and much larger—group with some of these same characteristics. These are the clients here designated as "test preoccupied." In the spontaneous comments offered, some 15 per cent of the group indicated, in one way or another, that the tests were the most valuable part of the service. That is to say, such people mention only this phase of VSC's work, either favorably or not; they never talk about benefits that might have accrued from counseling. It would appear that the sole benefit—if, indeed, the respondents would term it that—received from VSC was the time spent in the testing rooms. Not all of these are the "low achievers" as described in Friend and Haggard's monograph (5), but some of them are: people with poor emotional attitudes towards work, individuals with considerable rigidity as regards personality structure. Some, however, have already shown signs of good achievement. But the sheer frequency with which such comments crop up suggests client orientation toward vocational counseling that needs changing.

As Covner (3) and others have pointed out, the usefulness of aptitude tests where it is demanded that counselors provide specific occupational information has long been apparent. An approach that is not exclusively oriented to aptitude testing may easily be applied both to the preparation of the client for tests and to interpretation of such results. The comments of this group evidence the fact that many leave the entire counseling set-up with attitudes about tests that need redirecting.

Others blame the tests for lack of specificity. The tests only told them what they already knew; there were no tests available for some special occupational area, etc. A few respondents specifically refer to the tests, especially the interest inventories, as being too suspectible to "faking." Several respondents, as if not satisfied with the tests administered, want more. Possibly they have changed their occupational goal but still find themselves undecided and uncertain. In any event, they look to "the tests" for answers. The psychological insecurity of some of these clients is apparent from their comments. A more exaggerated picture of this is to be seen in clients who have "shopped around" for other tests. The most extreme case is the one respondent whose long letter (13 pages) indicated he had been to a minimum of seven different counseling agencies and had been given tests at all.

It should be borne in mind by the reader that such "test preoccupied" clients only represent a small proportion of the total group (15 per cent). A direct question about the test attitudes of respondents was not asked. Had this been so, a larger number of such replies might well have been received. Stone and Simos (11), for example, did ask such a question and found some 60 per cent with "test oriented" attitudes. Covner (3) found this a prevalent tendency in his clients; Kohn (7) at the University of Washington reports more than 60 per cent of veterans registering at the university's counseling center said they desired vocational tests.

At any rate, the counseling process has been something less than a success with many of this group who often need a discussion about unconscious emotional attitudes. Others seem to have only a mild interest in improving present circumstances. Small wonder these continue to look to "the tests" for the answers and continue to insist on greater and greater specificity both in test results and counselor suggestions. Any sound counseling approach which recognized motivation as the primary datum would be of great help here—i.e., one which would be organized around the discovery and correlation of the dynamics of the individual with those of the occupational objective that is to be approved.[3]

Questions Raised

From many veterans the receipt of the questionnaire brought an immediate response and one that entailed considerable correspondence. It was as if the receipt of the questionnaire revived many unanswered questions in this group and, having had one helpful experience, they turned again to the agency. Numerically

[3] See Part IV: "A Study of Job Values" in Friend and Haggard (5) for a conceptual basis upon which such a program might be built. McGregor (8) has discussed the problems of industrial leadership and supervision from the same point of view.

the group is not a large one, in the light of all returns, but the extra time involved with these cases was extensive. A total of 64 individuals (11 per cent of the comment group) raised questions demanding some sort of reply. Many desired a written summary of test results, which was duly sent out; many more betrayed feelings of insecurity about present status, some requesting another interview while others left it up to VSC. The breakdown is as follows: 20 (31 per cent) asked for test scores; 31 (48 per cent) indicated further counseling was needed; 8 (12 per cent) placement help was requested; 5 (9 per cent) miscellaneous.

By and large these individuals are the "test preoccupied" people. They are compulsive about the tests they have already taken; they want more in order to bolster feelings of self-confidence. They ask unrealistic questions about schooling or they want extra help in other ways. Some did not even wait to write a letter but came right in for a visit after receipt of the questionnaire.

Implications for the Counselor

What are the implications for the counselor? The comments show that clients have done considerable after-visit evaluation. From this welter of information some specific suggestions for counselors may be garnered. The respondents have provided a picture of the counseling process as seen from their vantage point. It is evident that the typical client arrives at this situation with two dominant sets, namely, that the counselor will tell him what is best and that the tests will provide most of the answers.

Few good counselors, of course, really play an authoritarian role. It would certainly be wise, with the more dependent clients at least, specifically and explicitly to delineate the counselor-client relationship as a cooperative one. The essential purpose of all sessions is clarification and not to inform the client that he is best equipped to pursue, say, bookkeeping versus accounting. It might be well to state at the outset that the arrival at *any* specific answer regarding future occupational goals is unlikely.

The notion that vocational and educational problems can be answered by tests must be altered. Here might well be applied a tactic so that the counselor secures a picture of what the client feels on this score at the beginning. There are, after all, vocational problems where it would be unnecessary to administer any tests at all. Interview manuals often refer to the "preparation of the client for tests." It is precisely at this stage where the counselor must so structure his remarks that false hopes and illusions about test results are not established in the client. Pursuing this tactic further, the client may frequently be allowed to participate in the actual business of deciding whether certain tests are suitable for his particular case; thus all responsibility is not placed on the counselor. Bordin and Bixler (**2**) provide excellent suggestions here; Seeman's study (**9**) on such a self-selection-test technique is also pertinent.

The interpretation of test results to the client frequently demands all the subtlety, tact and skill that a counselor may muster. It is suggested here that it might be a useful device to secure first from the client a "free" response as to how he thought he performed. By such a means the counselor might quickly tell whether he need proceed cautiously and with understatement or otherwise. Often such a "free" response will disclose that the client has already done some accurate self-evaluation which might now be woven into the counseling fabric. Covner (**3**) has dealt with this aspect; Bixler and Bixler (**1**) offer detailed quotations and suggestions on this interpretation problem. At this stage, too, the counselor might attempt to forestall unfavorable reactions (frequently met) with clients actually launched on a job or school career or who are going through such services only because of necessary VA approval. The type of interpretation of test data for this sort of client should be quite different from that given to an individual who has come for counseling because of real need.

Many clients are in need of a frank statement about the nature and purpose of interest inventories. The impression is to be avoided that interests, either claimed or measured, are of primary significance. This is frequently just the sort of notion many clients get, possibly because it is relatively easy for the

counselor-client relationship to get under way by first discussing likes and dislikes. At all costs, the final impression, prevalent among the critical respondents, that all the counselor did was to agree with originally stated likes of the client, is to be avoided. Similarly, too complete reliance on high points of interest profiles may be of doubtful validity, as Diamond (4) has recently discussed. In addition, a succinct caution is sometimes necessary about the function of interest inventories and about the possible operation of unconscious sets on the part of clients that might distort scores.

Counselors should take pains to explain clearly the type of norm data used on which raw scores are evaluated. Many veterans of beyond high school level are deflated when they learn they receive above average scores on tests where twelfth grade norm groups are the standard. This is part and parcel of the need of developing local norms.

Several respondents suggested that some sort of a written summary for each client be provided. The idea, on the surface, appears to be a good one. This might easily be a one-page sheet where test scores are indicated in general terms along with a brief description of the test itself. It is suggested that any precise indication of test scores—such as percentile ratings, sigma scores, etc.—be avoided. To the psychometrically unsophisticated person such devices tend to indicate greater accuracy than they possess.[4]

There must be an honest facing on the part of the counselor towards all client handicaps—such as a client's lack of a high school diploma. No sugarcoating, as clients are quick to perceive this type of "dishonesty"; they specifically mention they resent being sent away happy. Here also one might employ a non-directive question or two, so as to get the client to express in his own words what he feels his present handicaps to be. This would, in turn, help to minimize the "telling" on the part of the counselor.

[4] Such a practice contains hazards; it is not VSC policy. Here clients who desire further interpretation of test results come in for an additional appointment and the test results are then presented verbally. Such a practice obviously avoids the pitfalls inherent in any procedure that allows the client to draw his own conclusions.

As much advance planning should go into the summary interview as possible. Clients show by their comments that they arrive at this stage somewhat "keyed-up." Presentation of facts about test performance is important but is not the main function of counseling. Here is where skillful weaving of test result data into the counseling situation is paramount. If the client has been well prepared for the tests in preliminary interviews, he will not expect the counselor merely to sit back and "tell." This interview might well be the longest as regards actual time. Clients expect, often unwisely, great things here; they are disillusioned by speed, brevity or anything that might be interpreted as vagueness.

A closing suggestion from the counselor that he would be glad to talk again with the client at some future date would be very acceptable to many. To be sure, this is generally assumed—but often only by the agency personnel and is rarely stated explicitly. This then puts the responsibility for a follow-up squarely on the client. The spontaneous comments mention that clients are very receptive to even a mailed questionnaire one or more years later; even this check prompted several to call at the office for interviews.

It must be recognized at the outset that not all comers are fit for counseling, since there may be dominant unconscious emotional forces operating which vitiate all suggestions no matter how carefully thought out and handled by the counselor. A non-directive beginning with clients might well unearth some of these "unfavorables" near the starting line. The more that is known of these dynamics, the better will all vocational and educational counseling become.[5]

Summary

The reactions to counseling of 890 veterans were studied by means of replies to items on a mail questionnaire together with spontaneous comments. The counselor suggestions were found to be genuinely helpful by 75 per cent of the group; 91 per cent reported the staff

[5] Friend and Haggard have not only stated the problem explicitly but have made a significant beginning towards this goal; see especially pages 110–111 of their monograph (5).

attitude a sympathetic one. Counseling served to increase the self-confidence of many. The prevalence of a "test preoccupied" group is noted. Implications for the psychological counselor are discussed in terms of the comments offered by these veteran clients.

Received February 16, 1950.

References

1. Bixler, R. H., and Bixler, V. H. Test interpretation in vocational counseling. *Educ. psychol. Measmt.*, 1946, **6**, 145–156.
2. Bordin, E. S., and Bixler, R. H. Test selection: a process in counseling. *Educ. psychol. Measmt.*, 1946, **6**, 361–374.
3. Covner, B. J. Nondirective interviewing techniques in vocational counseling. *J. consult. Psychol.*, 1947, **11**, 70–73.
4. Diamond, S. The interpretation of interest profiles. *J. appl. Psychol.*, 1948, **32**, 512–520.
5. Friend, J. G., and Haggard, E. A. Work adjustment in relation to family background. *Appl. Psychol. Monogr.*, 1948, No. 16. Pp. 150.
6. Hawkins, L. S., and Fialkin, H. N. *Clients' opinions of the Adjustment Service.* New York: Amer. Assoc. Adult Educ., Adjustment Service Report XII, 1935. Pp. 95.
7. Kohn, N., Jr. Trends and development of the vocational and other interests of veterans at Washington University. *Educ. psychol. Measmt.*, 1947, **7**, 631–637.
8. McGregor, C. Conditions of effective leadership in the industrial organization. *J. consult. Psychol.*, 1944, **8**, 55–63.
9. Seeman, J. A study of client self-selection of tests in vocational counseling. *Educ. psychol. Measmt.*, 1948, **8**, 327–346.
10. Seipp, E. *A study of 100 clients of the Adjustment Service.* New York: Amer. Assoc. Adult Educ., Adjustment Service Report XI, 1935. Pp. 31.
11. Stone, C. H., and Simos, I. A follow-up study of personal counseling versus counseling by letter. *J. appl. Psychol.*, 1948, **32**, 408–414.

The Wechsler-Bellevue Intelligence Scale and High School Achievement *

Arden N. Frandsen

Utah State Agricultural College

The validity of the Wechsler-Bellevue Intelligence Scale as a measure of intelligence as indicated by its correlations with other already established tests of intelligence has been reported by several writers. Its correlation with various other tests have ranged from .39 to .93 (**25**, p. 134). The median of six different correlations with the Revised Stanford-Binet is .87 (**8**, **16**, **22**, **25**). And Altus and Mahler found a bi-serial *r* of .52 between W-B IQ's and graduation versus being discharged as inept from the 9th Service Command Special Training Center (**1**). Tests, however, have multiple validities, which may vary with differences in the particular criteria one may wish to predict. High school counselors may be especially interested in the validity of the W-B, both as a whole and when abbreviated by division into selected subtest groups, for predicting achievement in high school curricula. The data to be presented in this paper should contribute specifically to this question, and by implication, also to an evaluation of the test for college guidance.

Two previously reported investigations have contributed data on the validity of the W-B in predicting academic achievement at the college level, a situation closely related, of course, to the question of the test's validity for predicting high school academic achievement. Anderson (**2**), for 112 college women whose mean W-B IQ was 118.5, found correlations with average freshman grades of .45 for full scale IQ, .52 for verbal IQ, and of .23 for performance IQ. Sartain (**22**), for 50 college students whose mean IQ was 117.5, found the corresponding correlations of freshman grades with W-B IQ's to be .53, .58, and .35, respectively. These coefficients compare very favorably, especially for the W-B verbal scale, with the modal range of correlations (.40 to .60) which have been reported between high school grades and group verbal intelligence tests (**6**, p. 846).

Subjects. The subjects used in the studies conducted by Anderson and by Sartain were not selected to represent adequately the range of abilities of college freshmen, nor was the area in which achievement was appraised homogeneous. The subjects of the present study comprise practically an entire class (83 out of 90) of relatively superior high school seniors from a mid-west university city whose employed population consists predominantly of university staff and otherwise of people employed in professional and business occupations. A large proportion (83%) of the high school seniors anticipated attending college and were highly motivated to meet entrance requirements. There were variations in the courses pursued, but nearly all students included a basic core of college-preparatory subjects in their courses of study. And because these high school students were living with their own families, within a fairly narrow range of socio-economic status, their general conditions of living were also probably more similar than they were for the college freshmen. The ages (obtained when the W-B was given) of these 83 high school seniors ranged 16 to 19 with a mode of 17. The range in W-B IQ was from 86 to 145, and the mean was 119.8.

Hypotheses. In this study, the data on intelligence and academic achievement will be analyzed to test three hypotheses. 1. With conditions more favorable for predicting achievement (a more complete range of abilities represented, high motivation, more homogeneous curricula, and probably more uniform general conditions of living), validity coefficients even higher than those found by Anderson and by Sartain for college students may be obtained. 2. Because certain subtests may dilute the predictive efficiency of the W-B for a particular purpose, it will be possible, by omitting the subtests showing lowest correlations with our criterion, to find a combination of subtests that will predict this criterion more effectively than the present full scale. 3. As a corollary to

* To Mr. Carl Hammer, principal, and to Mrs. Gladys Issacs, Counselor, both of the West Lafayette, Indiana High School, the writer is very grateful for their interest and generous help in arranging for testing their students and for supplying the achievement data.

hypothesis 2, for testing situations where time economy is important, valid abbreviated W-B scales may be selected wisely according to their separate validities in predicting a particular criterion. Even with drastic deletions from the full scale, the greater saturation of the abbreviation with valid items (for a particular criterion) may compensate for the reduced reliability of the shorter tests.

The first hypothesis is similar to that proposed by Burtt and Arps to explain their higher correlation between Alpha intelligence and academic achievement in military academies compared to public high schools, *r*'s being .39 and .19, respectively. In the academies, they they observed, the "general supervision, supervised study, and a system of reward and punishment are more apt to hold the student to his maximum intellectual ability" (**3**, p. 393). The latter two hypotheses are rendered plausible by the facts that the 11 different subtests of the W-B were selected to test a variety of mental functions and that the 45 intercorrelations reported by Wechsler (**25**, p. 223) for 10 of the subtests indicate only a moderate degree of communality among the mental functions being measured by the scale. They range, except for only 11 coefficients outside this range, from .3 to .5, with the mode being .4. It is quite probable, therefore, that the separate tests are measuring to some extent different quasi-general aptitudes and/or specific abilities. And the comparability of the 11 W-B subtest scales makes it feasible and convenient to take advantage of this flexible feature of the W-B test.

Procedure. The data of this study include, for 83 high school seniors: (1) average grade-point-ratios earned in an academic curriculum and cumulated from grade 9 through 12; (2) Henmon-Nelson, Form A IQ's; and (3) Wechsler-Bellvue IQ's and separate subtest weighted scores. The W-B tests were administered to the high school students, at different periods during their senior year,[1] by the writer's Purdue University students in Individual Intelligence Testing, *after* they had demonstrated to the writer adequate skill in administration of the test. Product-moment correlations were determined between grade-point-ratios and the IQ's for both tests and with the W-B subtests, singly and in various combinations.

[1] Ideally in a predictive study the prognostic tests should have been administered at the beginning of the period covered; much data on IQ constancy, however, perhaps justify the assumption that IQ's obtained in the senior year would not have differed significantly from IQ's that should have been obtained two years earlier.

Results

1. *IQ's and Grade-Point-Ratios.* The data for interpreting IQ's and for evaluating their validity in predicting academic achievement in this high school are presented in Table 1.

Table 1

W-B IQ and Henmon-Nelson IQ Data for 83 High School Seniors and the Correlations of the Tests with Total (3 years) Grade-Point-Ratios

Tests	Means	S.D.	*r* with G.P.R.
Wechsler-Bellevue, F-IQ	119.8	11.86	.685
Wechsler-Bellevue, V-IQ	115.7	12.96	.69
Wechsler-Bellevue, P-IQ	119.6	10.08	.48
Henmon-Nelson, Form A IQ	112.1	12.84	.52

Taking as a basis of comparison the Henmon-Nelson IQ and grade-point-ratio correlation of .52, which is as high as has usually been reported in the literature (**6**, p. 845) for similar prognostic relationships, the correlations between W-B IQ's and grade-point-ratios are surprisingly high. As is shown in Table 1, the W-B verbal IQ alone and the full scale IQ are equally good predictors of high school academic achievement, the correlations being .69 and .685, respectively. For the performance scale, the correlation of .48 with grade-point-ratio is much lower, but it is only slightly lower than the corresponding correlation of .52 with the Henmon-Nelson group test IQ's. The relatively high validity of the W-B, especially the verbal scale, in predicting high school achievement, although perhaps surprising in comparison with the results previously reported both for group tests and for the W-B on the college level, is in line with our hypothesis that when the factors of motivation, and uniformity of curriculum and of living conditions are all favorable, higher validities are to be expected.

Referring again to Table 1, it should also be noted that the mean IQ of our group is relatively high, being 119.8 for the W-B full scale. Because of this fact and in the light of Goldfarb's contention that the W-B is relatively ineffective in discriminating among a group of superior adolescents (**8**), it might have been inferred that the test would have proved to be an inefficient guidance device for our population. The correlation results, however, are inconsistent with this assumption; and three other indices indicate adequate discrimination among our population of superior high school students: (1) the IQ's vary in range from 86 to 145, (2) the standard deviation of 11.86 is only 2.64 points narrower than the standard deviation of 14.50 for Wechsler's standardization group of comparable age (**25**, p. 122), and (3) the distribution of IQ's appears approximately normal, being not markedly bunched toward the highest IQ's. In interpreting IQ's for these superior students, however, local in addition to "national" norms are helpful. For example, an IQ of 120 which corresponds to a Wechsler percentile of 93 is the equivalent of a percentile score of only 52 according to the local norms. In relating IQ's to probable potential achievement within the high school, the local norms are more consistent; but for vocational guidance, where the scope of educational and occupational competition is national, percentiles based on Wechsler's more representative standardization group are, of course, more appropriate.

In passing, it may also be interesting to note the intercorrelations between the Henmon-Nelson and the Wechsler-Bellevue IQ's. That the W-B verbal scale and the H-N measure quite similar traits is indicated by the correlation between them of .80. Between the W-B full scale IQ's and the H-N IQ's the correlation is .70, indicating fair agreement. But the correlation between H-N IQ's and W-B performance IQ's of only .44 indicates that the latter scale is measuring traits somewhat independent from the verbal abilities measured by both the H-N and the W-B verbal tests.

2. *Correlations between Each W-B Subtest and Grade-Point-Ratio.* As a preliminary both to determining the combination of subtests having a maximum correlation with academic achievement and to selecting promising subtest combinations for abbreviated scales, the mean subtest weights and their correlations with average grade-point-ratios were determined for each of the 10 subtests administered to our population of 80 high school seniors.[2] These data are presented in Table 2.

In Table 2, attention is first directed to the variations in subtest means. Two notably low scores are Digit Span and Arithmetic, tests which have in common the fact that they both involve numbers and require for successful performance high degrees of attention (especially digits) and concentration. Findings such as those of Lewinski (**17**) together with the fairly frequent references in the test protocols to comments by the students on feelings of inadequacy on the "number tests" suggest that in several instances efficiency on these two tests may have been impaired for these highly motivated students by temporary conditioned emotional "panics." On the Block Designs test, which involves mainly analytic and synthetic abilities in dealing with abstract elements in multiple form-color relationships,

Table 2

Means and Correlations Between Separate W-B Subtest Weights and Academic Achievement (3 Years Cumulative Grade-Point-Ratios) for 80 High School Seniors

	Mean* Wt.	r
Verbal Tests		
Information	11.9	.56
Comprehension	13.0	.51
Digit Span	10.3	.45
Arithmetic	10.6	.55
Similarities	12.5	.54
Performance Tests		
P. Arrangement	12.5	.15
P. Completion	12.7	.26
Block Design	14.2	.44
Object Assembly	13.3	.11
Digit Symbol	12.6	.36

* In computing means, lower and upper limits of score of 10, say, were assumed at 9.5 and 10.5, respectively.

[2] The number of subjects is reduced from 83 to 80 because for three students complete separate subtest scores were unavailable.

these superior adolescents achieved their best performance.

More important as a guide in selecting valid subtest combinations, however, are the correlations of the separate subtests with the achievement criterion. On the whole, as might have been expected from coefficients reported in Table 1, the verbal tests are distinctly more valid for this purpose than are the performance tests, with the exception of the Block Designs test. It may be interesting to note that among the relatively high academic achievers, none earned a low Block Design score. However, several of the low achievers earned inconsistently high Block Design scores. It would seem that the kind of "reasoning" measured by this test may be necessary for high academic achievement, but it alone does not guarantee it. For three subtests (Picture Arrangement, Picture Completion, and Object Assembly) the correlations with grade-point-ratios fail to meet the statistical criterion for significance at the 1 per cent level of confidence. The scatter diagram for PC, on which test the ceiling limits weighted scores to 15, indicates that for these subjects this subtest is not discriminative; scores regardless of academic achievement tended to approach the ceiling. For both PA and OA, although the scores were well distributed over the full range, they apparently corresponded with achievement at no part of the range.

3. *Most Valid W-B Subtest Combination.* According to our second hypothesis, W-B subtests which are separately invalid for predicting the criterion may dilute the predictive efficiency of the other more valid subtests in the full scale. Table 2 clearly indicates three such relatively invalid subtests, the three—PA, PC, and OA—have already been mentioned as having statistically insignificant correlations with grade-point-ratios. Moreover, although all of the W-B subtests correlate significantly (.41 to .73) with the full scale, it happens that, except for Digit Span, these three tests also have the lowest correlations (.41 to .61) with the scale as a whole (**25**, p. 224). Therefore, when these tests are deleted, the remaining tests should have a greater saturation both with whatever is required to predict our criterion (academic achievement) and with the traits which the scale as a whole are measuring. The combination 10/7 (I + C + D + A + B + Dsy), which includes all of the subtests except PA, PC, and OA, was found to correlate .765 with grade-point-ratios, and thus supports our hypothesis quite well. It is recognized, of course, that we have not demonstrated statistically that this combination is *the* combination of maximum validity. Table 3, however, shows that compared to either the full scale or to any abbreviated combination which was evaluated, this combination predicts most efficiently our criterion.

4. *Validity of Abbreviated Scales.* To meet the need for an individual intelligence examination in situations where economy of time is important, several abbreviations of the W-B have been proposed and evaluated (**4**, **7**, **9**, **10**, **11**, **12**, **13**, **14**, **15**, **18**, **19**, **21**, **23**, **24**). In these explorations to find satisfactory abbreviations, the usual requirements in addition to economy of time have been: (1) mean of abbreviation approximating closely the full scale mean; (2) abbreviation IQ's correlating well with full scale IQ's; and (3) combinations of subtests likely to yield data which may be significant in the clinical diagnosis of personality. Criterion 1 has usually been satisfactorily attained by simple pro-rating of test weights or as in two instances (**4**, **15**) by special factor weightings. Correlations ranging from .80 to .97 indicate that criterion 2 has been acceptably met. Except in one investigation (**11**), however, wherein abbreviation "CVS" subtest pattern was found to distinguish statistically both aged and schizophrenic subjects from normal and feeble-minded subjects, the claims that the abbreviations would yield significant "clinical" data have been supported only on the basis of logic and clinical experience. In this study of W-B abbreviations we shall be concerned with another criterion, namely their validity in predicting academic achievement in high school. In selecting promising subtest combinations, we shall, therefore, besides considering the abbreviations already proposed[2] and the "aptitude" groupings suggested by

[2] Because we did not regularly administer the Vocabulary subtest, some abbreviations (**10**, **11**, **12**, **18**, **19**) which included this excellent test or an abbreviated substitute for it (**14**, **24**) were omitted in our evaluations.

Table 3

Mean IQ's and Correlations with Academic Achievement of the W-B Full Scale IQ, the Most Valid W-B Combination, and 12 Selected W-B Abbreviations for 80 High School Seniors

Combination of Subtests	IQ for Mean Pro-Rated Wts.	r with G.P.R.
a. W-B full scale IQ	119.8	.685
b. 10/7 (I+C+D+A+S+B+Dsy), Highest validity	118.0	.765
1. Geil's 5/2 (C+S+D+B)	121.7	.71
2. Patterson's "Reasoning Group" 5/2 (A+S+B+C)	123	.69
3. Rabin's 10/3 (C+A+S)	118	.69
4. 10/3 (I+A+S), Highest 3 r's with G.P.R.	114	.69
5. Diamond's "Linguistic Factor" 10/3 (I+C+S)	119	.65
6. Diamond's "Clerical Factor" 10/3 (D+A+Dsy)	118.5	.52
7. Diamond's "Spatial Factor" 10/3 (PC+OA+B)	112	.37
8. 10/3 (C+A+B), Highest r's plus qualitative clues	121	.62
9. Cummings, MacPhee, & Wright's 5 (C+A)	116	.625
10. Gurvitz's 5 (PA+D)	112	.39
11. 5/2 (I+A), Highest 2 r's with G.P.R.	111	.63
12. 5 (A+B), Reasoning: Numerical and spatial symbols	120	.60

the factor research of Diamond (**5**) and Patterson (**20**), be guided by Table 2 showing the correlations of separate subtests with our criterion. All of these combinations and the data for evaluating them are presented in Table 3.

The data in Table 3 support quite well our third hypothesis, that concerning the possibility of finding W-B abbreviations which are valid for a particular purpose. According to the criterion of validity in predicting academic achievement, the table shows that four abbreviations—numbers 1, 2, 3, and 4—equal the efficiency in this respect of the full scale, although they require only about half the time for administration. Five other three-test or even two-test abbreviations—numbers 5, 8, 9, 11 and 12—besides being even more economical of time, are reasonably efficient in predicting academic achievement, the correlations ranging from .60 to .65. All are superior in this respect to the Henmon-Nelson group test of intelligence. That the abbreviations are not all effective, however, is clearly indicated by the low correlations of .37 and .39 for Diamond's "Spatial Factor" and Gurvitz's "ten-minute" scale. Nevertheless, three two-test abbreviations—numbers 9, 11, and 12—predict academic achievement for these students fairly efficiently, r's ranging from .60 to .63.

Summary

Using as subjects practically an entire class of 83 relatively superior high school seniors, the validity of the Wechsler-Bellevue Intelligence Scale, both as a whole and in various subtest combinations, has been evaluated for effectiveness in predicting academic achievement.

1. Both W-B full scale IQ's and verbal scale IQ's are found to predict three year average grade-point-ratios very efficiently, r being .69 for both tests. The W-B performance scale, however, measures traits less significantly related to academic achievement.

2. When the subtests found separately least valid for predicting achievement are deleted from the full scale, the resulting combination, 10/7 (I + C + D + A + S + B + Dsy), is found to correlate even higher (.765) with grade-point-ratios. This is probably the W-B combination of maximum validity in predicting high school academic achievement.

3. By choosing subtest combinations according to their correlations separately with the criterion, at least nine different W-B abbreviations of satisfactory validity are found, correlations with academic achievement ranging from .60 to .71.

Our data are limited to high school seniors,

but in the light of the results obtained by Anderson (2) and by Sartain (22), it is probable that these conclusions would also apply with college freshmen, when conditions equally favorable to effective prediction are present.

Received January 12, 1950.

References

1. Altus, W. D., and Mahler, C. A. The significance of verbal aptitudes in the type of occupation pursued by illiterates. *J. appl. Psychol.*, 1946, **30**, 155–160.
2. Anderson, E. E., and others (Wilson College Studies in Psychology). A comparison of the Wechsler-Bellevue, Revised Stanford-Binet, and Am. Council on Educ. Tests at the college level. *J. Psychol.*, 1942, **14**, 317–326.
3. Burtt, H. E., and Arps, G. F. Correlation of Army Alpha Intelligence Tests with academic grades in high schools and military academies. *J. appl. Psychol.*, 1920, **4**, 289–293.
4. Cummings, S. B., Jr., MacPhee, H. M., and Wright, H. F. A rapid method of estimating the IQ's of subnormal white adults. *J. Psychol.*, 1946, **21**, 81–89.
5. Diamond, S. The Wechsler-Bellevue Intelligence Scales and certain vocational aptitude tests. *J. Psychol.*, 1947, **24**, 279–282.
6. Eurich, A. C., and Cain, L. R. Prognosis. In Monroe, W. S. (Ed.), *Encyclopedia of educational research*. New York: Macmillan, 1940, pp. 838–850.
7. Geil, G. A. A clinically useful abbreviated Wechsler-Bellevue Scale. *J. Psychol.*, 1945, **20**, 101–108.
8. Goldfarb, W. Adolescent performance in the Wechsler-Bellevue Intelligence Scales and the Revised Stanford-Binet Examination, Form L. *J. educ. Psychol.*, 1944, **35**, 503–507.
9. Gurvitz, M. S. An alternate short form of the Wechsler-Bellevue Test. *Amer. J. Orthopsychiat.*, 1945, **15**, 727–733.
10. Hunt, W. A., Klebanoff, S. G., Mensh, I. N., and Williams, M. The validity of some abbreviated individual intelligence scales. *J. consult. Psychol.*, 1948, **12**, 48–52.
11. Hunt, W. A., French, E. G., Klebanoff, S. G., Mensh, I. N., and Williams, M. The clinical possibilities of an abbreviated individual intelligence test. *J. consult. Psychol.*, 1948, **12**, 171–173.
12. Hunt, W. A., French, E. G., Klebanoff, S. G., Mensh, I. N., and Williams, M. Further standardization of the CVS Individual Intelligence Scale. *J. consult. Psychol.*, 1948, **12**, 355–359.
13. Hunt, W. A., and French, E. G. Some abbreviated individual intelligence scales containing nonverbal items. *J. consult. Psychol.*, 1949, **13**, 119–123.
14. Hunt, W. A., and French, E. G. A second fifteen-word vocabulary test for use with abbreviated intelligence scales. *J. consult. Psychol.*, 1949, **13**, 124–126.
15. Kriegman, G., and Hansen, F. W. VIBS: A short form of the Wechsler-Bellevue Intelligence Scale. *J. clin. Psychol.*, 1947, **3**, 209–216.
16. Kutash, S. B. A comparison of the Wechsler-Bellevue and the Revised Stanford-Binet Scales for adult defective delinquents. *Psychiat. Quart.*, 1945, **19**, 677–685.
17. Lewinski, R. J. The psychometric pattern: I. Anxiety neurosis. *J. clin. Psychol.*, 1945, **1**, 214–221.
18. Patterson, C. H. A comparison of various 'short forms' of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1946, **10**, 260–267.
19. Patterson, C. H. A further study of two short forms of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1948, **12**, 147–152.
20. Patterson, C. H. *Using the Wechsler-Bellevue Intelligence Scales*. In-Service Training Section, Personnel Div., VA Regional Office, Minneapolis, Minn., 1948.
21. Rabin, A. I. A short form of the Wechsler-Bellevue Test. *J. appl. Psychol.*, 1943, **27**, 320–324.
22. Sartain, A. O. A comparison of the New Revised Stanford-Binet, the Bellevue Scale, and certain group tests of intelligence. *J. soc. Psychol.*, 1946, **23**, 237–239.
23. Springer, N. N. A short form of the Wechsler-Bellevue Intelligence Test as applied to the naval personnel. *Am. J. Orthopsychiat.*, 1946, **16**, 341–344.
24. Thorndike, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26**, 128–135.
25. Wechsler, D. *The measurement of adult intelligence*. Baltimore, Williams and Wilkins Co., 1944.

The Minnesota Clerical Test: Sex Differences and Norms for College Groups

Olga E. de Cillis Engelhardt

Department of Psychology, The University of Connecticut

The Minnesota Clerical Test is extensively used in business and industry (**5**), in vocational advisement (**3**), as a laboratory exercise in the introductory course in psychology (**6**), and to predict achievement in selected college courses (**4**). The manual for this test is replete with 54 different sets of norms, none of which, however, are for college students.

In connection with a larger research project being conducted by the School of Business Administration at The University of Connecticut, the author had occasion[1] to administer the Minnesota Clerical Test to large groups of college students. The purpose of this brief paper is two-fold: (1) to present a needed set of norms for college students, and (2) to confirm for college students the finding of a sex difference in performance previously reported by Loevinger (**2**), and Thatcher (**2**) for grade school children, by Schneidler (**2**) for high school children, and by Andrew and Paterson (**1**) and Schneidler and Paterson (**7**) for adults in the general population.

Table 1

Norms for Eastern Arts and Sciences College Students

	Women		Men	
Centile	Test 1 Numbers	Test 2 Names	Test 1 Numbers	Test 2 Names
99	187	189	166	172
90	164	177	147	151
80	148	164	137	144
70	139	155	131	136
60	133	144	124	128
50	128	140	116	120
40	121	131	109	115
30	115	123	105	104
20	107	108	98	97
10	100	93	88	84
1	80	77	75	67
N	101	101	141	141

Subjects and Procedure

The subjects of this investigation were 512 junior and senior undergraduate students at The University of Connecticut. Of these 141 male students and 101 female students were enrolled in the College of Arts and Sciences in advanced courses in English, government, psychology, and zoology. The Arts and Science men ranged in age from 20 to 28; the median age was 21. The women ranged in age from 19 to 26; the median age was 20. The School of Business Administration men consisted of 270 enrolled in a course in industrial psychology. They were primarily industry and marketing majors. These students ranged in age from 20 to 32; the median age of this group was 24. The Minnesota Clerical Test was administered

Table 2

Norms for Eastern Business Administration College Students *

	Men	
Centile	Test 1 Numbers	Test 2 Names
99	185	192
90	153	160
80	136	146
70	130	138
60	124	130
50	118	123
40	115	117
30	110	110
20	105	103
10	95	95
1	80	77
N	270	270

* There is not a sufficient number of women enrolled in the School of Business Administration to warrant the publication of norms.

[1] Graciously provided by Dean Lawrence J. Ackerman.

Table 3

A Comparison of the Medians for the College Groups and Those Reported in the Minnesota Clerical Test Manual*

Groups	Women		Men	
	Test 1 Numbers	Test 2 Names	Test 1 Numbers	Test 2 Names
Arts and Science Students	128	140	116	120
Business Administration Students	—	—	118	123
Eastern Clerical Applicants	128	132	114	121
Southern Clerical Applicants	103	102	—	—
Adults Gainfully Occupied	109	111	83	78
Employed Clerical Workers	144	152	135	126

* (2, p. 8.)

Table 4

Means and Standard Deviations of the Minnesota Clerical Test Scores

Test	N	Mean Raw Scores	Standard Deviations
Numbers			
Arts and Sciences Women	101	128.63	25.02
Arts and Sciences Men	141	117.32	21.95
Business Administration Men	270	121.52	21.74
Names			
Arts and Sciences Women	101	136.70	29.35
Arts and Sciences Men	141	119.66	25.45
Business Administration Men	270	125.44	25.17

Table 5

A Comparison of the Means and Standard Deviations of the College Groups with the General Population and the General Clerical Population*

Test 1, Numbers

	Gen. Population		Gen. Clerical		A.S. College		B.A.
	Men	Women	Men	Women	Men	Women	Men
M	83.1	113.1	136.1	138.6	117.32	128.63	121.52
σ	29.2	30.8	27.9	26.3	21.95	25.02	21.74
N	491	229	41	59	141	101	270

Test 2, Names

	Gen. Population		Gen. Clerical		A.S. College		B.A.
	Men	Women	Men	Women	Men	Women	Men
M	77.6	111.3	130.0	144.4	119.66	136.70	125.44
	34.3	36.5	29.4	28.9	25.45	29.35	25.17
N	491	229	41	59	141	101	270

* (2, p. 304.)

during regular class periods during the academic year 1949–1950.

Results

Norms. The normative data for the college population are presented in Tables 1 and 2. A comparison of the medians obtained in the present study with those reported in the manual (**2**, p. 8) is made in Table 3. The college population medians for both tests are above those reported for adults gainfully employed and southern clerical applicants, but are below the medians for employed clerical workers. The medians of this study are most comparable to those reported for eastern clerical applicants.

Mean Raw Scores and Standard Deviations. The means and standard deviations for both tests for the college population are presented in Table 4. A comparison of the general population and general clerical population with the sample of this study is made in Table 5. College men and women obtain a higher average score than adults in the general population but a lower score than the general clerical population. With the exception of the standard deviation obtained in the present study, for women on test 2, all other measures of dispersion are smaller for the college group than for the comparison groups (Table 5). This is to be expected since the college population is more homogeneous.

Differences between Men and Women. The differences and their significances for both tests are shown in Table 6. All but one *t* value is significant at beyond the .1% level. The difference between Business Administration men and Arts and Sciences women is significant between the 1% and 2% level of confidence. The sex differences obtained in this study, although highly significant, are not as large as those reported in the literature (**1**, **7**). The significant differences between college men and women found in this study, however, substantiate the earlier findings.

Table 6

Differences between College Men and Women on the Minnesota Clerical Test

Test	Means	Mean Diff.	Diff.	*t*
Numbers				
A.S. Men	117.32	11.31	3.10	3.65**
A.S. Women	128.63			
B.A. Men	121.52	7.11	2.82	2.52*
A.S. Women	128.63			
Names				
A.S. Men	119.66	17.04	3.62	4.71**
A.S. Women	136.70			
B.A. Men	125.44	11.26	3.30	3.41**
A.S. Women	136.70			

* Significant between the one and two per cent levels of confidence.

** Significant beyond the one per cent level of confidence.

Summary

One hundred and forty-one male and 101 female undergraduate students in the college of Arts and Sciences and 270 males in the School of Business Administration at The University of Connecticut took the Minnesota Clerical Test. Sets of norms for these groups are presented. The sex difference as indicated by other studies on grade and high school children and adults is extended to include college students.

Received August 14, 1950.
Early publication.

References

1. Andrew, D. M., and Paterson, D. G. Measured characteristics of clerical workers. *Bull. Empl. Stab. Res. Inst.*, Univ. of Minnesota, 1934, **3**, No. 1.
2. Andrew, D. M., and Paterson, D. G. *Minnesota Clerical Test: 1946 revised manual for the Minnesota Vocational Test for Clerical Workers.* New York: The Psychological Corp., 1946.
3. Baker, G., and Peatman, J. G. Tests used in Veterans Administration Advisement Units. *Amer. Psychologist*, 1947, **2**, 99–102.
4. Barrett, D. M. Prediction of achievement in typewriting and stenography in a liberal arts college. *J. appl. Psychol.*, 1946, **30**, 624–630.
5. Bennett, G. K., and Wesman, A. G. Industrial test norms for a southern plant population. *J. appl. Psychol.*, 1947, **31**, 241–247.
6. Orbison, W. D., Bousfield, W. A., and de Cillis, O. E. *Group demonstrations and exercises in psychology.* Edwards Brothers, Inc., Michigan, 1950, 196.
7. Schneidler, G. G., and Paterson, D. G. Sex differences in clerical aptitude. *J. educ. Psychol.*, 1942, **33**, 303–309.

An IBM Card Profile for the Strong Vocational Interest Blank

Wilbur L. Layton

Student Counseling Bureau, University of Minnesota

In the winter of 1949 the Strong Vocational Interest Blank was administered to approximately 7,500 high school seniors in the state of Minnesota. This testing was a part of the state-wide high school testing program sponsored by the Association of Minnesota Colleges. Such large scale use of the Strong Blank was made possible by the cooperation of Mr. Elmer Hankes of Engineers Northwest.[1]

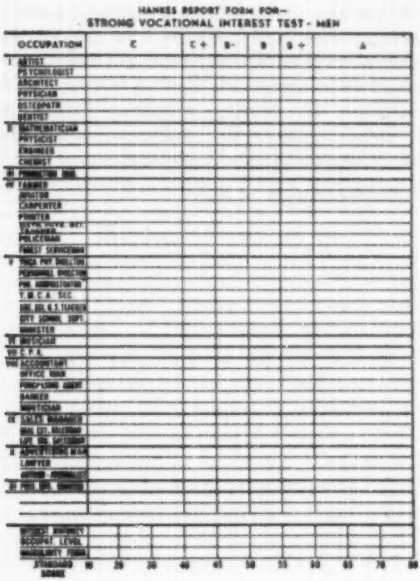


FIG. 1. Sample card for reporting Strong Vocational Interest results for men.

In order to eliminate the expense of manual duplication and filing of the profiles resulting from this program, the author designed two cards, one for men, one for women.

The cards were designed in a way such that holes punched in the front of the cards would indicate the letter grades on the various occupational scales printed on the back. In addition, the standard scores were coded so that frequency distributions and hence, statistics, could be obtained for each occupational scale.

The code system developed for each occupational scale on the Hankes report forms for men and women is shown in Table 1.

It will be seen that the code numbers 3 and 4, 5 and 6 can be combined to form class intervals of ten standard score units each. This gives seven class intervals of ten standard

Table 1

Transmutation Table for Occupational Scales

IBM Code Number	Standard Score	Strong Letter Rating
9	65–74	A
8	55–64	A
7	45–54	A
6	40–44	B+
5	35–39	B
4	30–34	B−
3	25–29	C+
2	15–24	C
1	5–14	C
0	−10–+4	C

Table 2

Transmutation Table for O. L., I. M. and M. F. Scales on Men's Form

IBM Code Number	Standard Score
9	70–79
8	65–69
7	60–65
6	55–59
5	50–54
4	45–49
3	40–44
2	30–39
1	20–29
0	10–19

[1] Strong, E. K., Jr., and Hankes, E. J. A note on the Hankes Test Scoring Machine. *J. appl. Psychol.*, 1947, **31**, 212–214.

Table 3

Transmutation Table for M. F. Scale on Women's Form

IBM Code Number	Standard Score
9	20–29
8	30–34
7	35–39
6	40–44
5	45–49
4	50–54
3	55–59
2	60–64
1	65–69
0	70–79

score units each and one class interval of 15 standard score units. Since standard scores below a minus five occur very rarely only a little error is introduced by assuming that one has eight class intervals of ten units each so that means, standard deviations and other statistics can be computed from the distribution obtained from the IBM cards.

The code for Occupational Level, Interest Maturity, and Masculinity-Femininity scales on the form for men is shown in Table 2.

Code numbers 3 and 4, 5 and 6, and 7 and 8 can be combined to give a total of seven class intervals of ten standard scores each.

The code used for the Masculinity-Femininity scale on the form for women is shown in Table 3.

Code numbers 1 and 2, 3 and 4, 5 and 6, and 7 and 8 can be combined to give a total of six class intervals of ten standard scores each.

This compromise between precise statistical coding and that needed for counseling seems adequate for most purposes. In the near future a revision of these forms will include several new scales which Hankes is planning to add to his present services. Shaded or "chance" areas on the various scales will also be incorporated in the revision.

Received February 16, 1950.

The Super-Roper Technique as a Measure of Interest in Nursing

Leslie Navran

Stanford University

In 1939, a systematic program of research was begun at Clark University under the direction of Donald Super (**1, 4, 5, 7, 8**) which was designed to explore the possibilities of an objective technique of interest measurement which involved a motion picture presentation of interests. Super's rationale was that a real interest in a vocational field would be manifested via attention to the film, and the closeness of attention and hence the amount of interest could be measured by a post-film test for recognition of what had been seen.

This paper is concerned primarily with the initial study, reported by Super and Roper (**8**). They showed a strip film (**6**) depicting the activities involved in nursing, as well as giving information about opportunities in the field, to 35 nurses and 111 high school senior girls, 33 of whom expressed a desire to go into nursing. The subjects were not told in advance of the post-film test. The Strong Vocational Interest Blank and the Otis S-A Test of Mental Ability (Higher Form, A) were also administered. For both nurses and students, the correlations between the Super Test of Interest and each of the other tests were zero. The interpretation suggested was that the Super Test was measuring something other than intelligence or subjective interest patterns. The authors argued that it was logical, therefore, to assume that the Super was a measure of interest in the occupation itself. The odd-even reliability of the test, corrected, was .83. Further analysis revealed that Super's test differentiated the nurses from the students, and within the latter group, those who wanted to be nurses from those who did not.

Older (**4**) has summarized the studies which were subsequently carried out. It need only be noted here that Bernstein (**1**) reported a correlation in the low .30's between the Super-type test for the metal trades and the Otis, when subjects expected a post-film test, and that Super and Haddad (**7**) found that Super-type test results were not significantly correlated with familiarity with or information about the occupations covered by the film. Older's (**4**) study has been the last reported to date. His results, based on a film which covered not one, but a range of occupations, were similar, in many respects, to those reported by Super and Roper.

The Problem

While Super's goal was to achieve a test which would be occupationally comprehensive, an equally valuable goal would be the further development of such a technique for a given specific field. Roper's (**5**) results raised the questions of: (1) the usefulness of such a technique in the selection of students for nursing; and (2) the relation between a high interest "in the work itself" and achievement. She made an attempt to investigate these questions, but the negative results were inconclusive, since the criteria were considered defective, and the sample included only *successful* trainees. This study was an attempt to make a more thorough investigation of the above questions. The focus of interest was not on the Super Test of Interest in Nursing per se, but in the technique and its potentialities.

Procedure

The subjects were 70 girls entering nurses' training in September 1947. Forty-four were divided among three hospitals affiliated with San Jose State College, and the remaining 26 were enrolled at a hospital affiliated with the Stanford University School of Nursing.

A Personal Data Sheet and an Information Blank, both of which were devised by the writer for this study, were added to Roper's (**5**) test battery. The Information Blank consisted of 15 items which formed a continuum of nursing knowledge ranging from that which is commonly known to that which is more likely to be known only by nurses. The test was pre-tested on a nursing sample and on a sample of senior high school girls not interested in nursing as a career.

In order to motivate the subjects, the battery was administered as a part of their orientation program. The order of the tests was as follows: Information Blank, film and Super Test, Personal Data Sheet, Otis, and Strong. Except for the new procedures, this was the order used by Roper. All the girls were free from other commitments for at least one hour before testing. Ventilation, visibility of the screen, and other conditions were as optimal as possible.

Two changes were made with respect to the film: (1) one of the film's 40 frames was exposed for 17 seconds instead of the usual 15, to insure its being completely read; and (2) the subjects were instructed to expect a post-film test which would be based on the film. This has been acknowledged by Super as a practical necessity.[1] With these exceptions, the comparison made later with Roper's data is between data obtained under essentially the same conditions.

The scoring of the Otis and Strong was in accordance with their respective manuals. The scoring key for the Super Test was supplied in Roper's (5) thesis, and verified by both Roper and Super.

The following two criteria of achievement were selected: (1) grades in courses taken during the first year of training, and (2) ratings of performance of duties while on the ward. Test results were withheld from the hospital and school administrators until the data relevant to the criteria were obtained, in order to avoid contamination of their judgements.

Results

In the first two years of training, ten girls left the San Jose group of trainees, four to be married and six because of scholastic difficulties. One girl left the Stanford class to be married.

Comparison of the two student groups' Personal Data Sheets yielded these major differences: (1) The Stanford girls were a more homogeneous age group and were two years older in median age. (2) Eighty-seven per cent of the San Jose girls had one year or less of college education, in contrast to the one Stanford girl (four per cent) who had less than two college years. This was a significant difference (C.R. = 7.0 $p; < .001$). (3) Only 52 per cent of the 19 San Jose girls who had some college education had major subjects related to nursing. In contrast, 88 per cent of the Stanford girls were nursing majors (C.R. = 2.6; $p < .01$).

The percentage of the girls in each group having work experience and previous social contact with medical persons was approximately the same. For both groups, the decision to enter nursing was based on a matrix of main reasons: a long-term ambition to be a nurse, a desire to help and work with people, an assortment of practical reasons, and a liking for the work. The most frequently given reason was: discussion with, and urging by, friends, family and counselors. However, only 12 of the 70 girls reported having had psychometric guidance.

The Information Blank significantly differentiated all the groups to which it was administered (see Table 1).

Going from the non-nurses to the trainees and nurses, there is a significant increase in the mean scores, a consistent drop in the variabliity, and a narrowing of the range. These results are what one would expect if the test were valid. The test also significantly discriminated the more highly educated and presumably better informed about nursing Stanford group from the San Jose group (C.R. = 3.75; $p < .001$). Although it may be argued that the small number of items did not

Table 1

Summary of Information Blank Scores

	H. S. Seniors (N = 82)	Nursing Trainees (N = 70)	Nurses (N = 51)
Mean	5.7	10.2	12.7
S.D.	2.45	2.2	1.4
Range	1–12	5–15	8–15

Critical Ratios

Nurses × Trainees	7.8 ($p < .001$)
Trainees × H. S. Girls	10.7 ($p < .001$)
Nurses × H. S. Girls	20.8 ($p < .001$)

[1] Personal communication.

Table 2

A Comparison of the Most Similar Components of Roper's and Navran's Sample with Respect to the Otis, Strong, and Super Tests

Test		Roper H. S. Girls Indicating Choice of Nursing (N = 33)	Navran H. S. Graduates Entering Nurses' Training (N = 44)	Critical Ratio
Otis	Mean	108.9	110.8	.90
	S.D.	9.9	7.95	
Strong	Mean	57.1	40.1	5.57
	S.D.	6.5	18.7	($p < .001$)
Super	Mean	51.1	87.5	12.59
	S.D.	12.0	13.2	($p < .001$)

tap all possible areas of nursing information, the obtained results indicate that these 15 items are a valid sampling of nursing information, which successfully differentiates better informed from less informed groups.

The mean Otis IQ of the Stanford group was 114.3, while that for the San Jose group was 110.8 (C.R. = 2.1; $p < .04$). Only the Strong failed to discriminate the two groups. The Stanford mean was 43.8 and the San Jose mean 40.1.

Super Test. For the purpose of investigating the test's reliability, all 70 subjects were grouped together. The correlation of the odd versus even items for the whole test was .54, raised to .70 by the Spearman-Brown formula. The first four parts of the test correlated .97 with the entire test. This supports Roper's (5) view that Part V is virtually useless. The odd-even reliability coefficient of these four parts was .48, raised to .65. The first three parts, reported by Super and Roper (8) as the most valuable for discriminating nurses from non-nurses, had a reliability of .52, raised to .69.

Comparison of Roper's Results with Those of the Present Study. As the San Jose group was closest to the 33 high school seniors in Roper's sample with respect to educational level, only the scores of these two groups were compared. This comparison is of interest, despite the San Jose group's being more highly selected (see Table 2).

The two groups were not significantly different in intelligence, as measured by the Otis. However, the two groups differed significantly on both the Strong and Super tests, Roper's sample having a marked edge on the Strong, while the subjects used in this study were superior on the Super.

The Strong Test comparison is clouded by the fact that a new revision of the test was used in this study. Ten items of the older form were cut out in this revision. Accordingly, some of the difference may be an artifact attributable to the Roper group's having a greater number of scoring opportunities. Probably of greater importance is the likelihood that the two groups, being seven years apart, actually differ with respect to the nursing pattern of likes and dislikes.

The Super Test difference may be accounted for by three factors: the extremely low reliability of the test, the fact that the subjects in this study expected the test, and the difference in selection.

Test Intercorrelations. A comparison of the interrelationships among the tests with the intercorrelations obtained by Roper (5) is presented in Table 3.

The correlation between the Super and Otis tests is the only one noticeably changed; the r of .345 is significant. This increase is in accord with the results reported by Bernstein (1) and Older (4).

Table 3

A Comparison of the Obtained Test Intercorrelations with Those Reported by Roper

	Navran (N = 70)	Roper Nurses (N = 35)	Roper H. S. Girls (N = 111)
Super × Otis	.345 ($p < .004$)	.006	.14
Super × Strong	.055	−.008	.001
Strong × Otis	.016	−.04	.00
Information × Otis	.395 ($p < .004$)	—	—
Information × Super	.198	—	—
Information × Strong	.001	—	—

The relation between information and intelligence, as measured by these tests, was positive and significant. The Super and Information Blank showed a positive but insignificant relationship. Evidently, the two tests are measuring essentially different functions. These results agree with those of Haddad and Super (7).

Comparison of Successful and Unsuccessful San Jose Trainees. An analysis of the test performance of the ten girls who left training, as compared to the 34 who remained, revealed no significant differences. However, the evidence indicated that those who left to get married were students as promising as those still in training. The scholastic failures scored lower than the marrying girls on each of the four tests, but only the difference in intelligence was significant ($t = 2.32$. 8 d.f.; $p < .05$).

Table 4

The Correlation of Tests in the Battery with the Criterion of Grades

Test	San Jose (N = 34)	Stanford-Lane (N = 25)	Critical Ratio
Otis	.33 ($p < .06$)	.65 ($p < .001$)	1.55
Information	.45 ($p < .01$)	.28	
Super	.20	.12	
Strong	.03	−.04	

The Relation Between Test Scores and the Criterion of Grades. For each student group, test scores were correlated with the average number of grade points for courses during the first school year (see Table 4). The 12 grades from A to F were equated with numerical values from zero to eleven, and the usual formula for computing grade (honor) points was used. The grade point average was obtained by dividing the total number of grade points by the total number of course units.

Neither interest measure correlated significantly with the criterion. The Otis correlations were low to moderately positive, and it would be the most useful test in the battery for prediction. Only in the San Jose group was the correlation for the Information Blank significantly positive. This suggests the possibility that with increased intelligence the function of previous information is minimized in mastering course work.

Table 5

The Correlation of the Parts of the Super Test with the Criterion of Grades

	San Jose (N = 34)	Stanford-Lane (N = 25)
Part I	.41 ($p < .02$)	.61 ($p < .01$)
Part II	−.20	.10
Part III	−.004	.12
Part IV	.35	.02
Parts I, II, III	−.002	.28
Total Test	.20	.12

The relationship of the various parts of the Super to the criterion is shown in Table 5.

Only Part I, having 21 multiple-choice items, was significantly related to the criterion in each group. The variability of the correlations of the other parts, both in amount and direction, was evidence of both their unreliability and their lack of relationship to the criterion.

The Relation Between Test Scores and the Criterion of Ratings of Performance of Duty in the Wards. Only the ratings for the Santa Clara County Hospital and Stanford Hospital trainees were suitable for analysis (see Table 6).

The correlations for the Santa Clara group are all lower than the Stanford correlations, but the rank order is the same in both groups. The Super Test correlations are not only low,

Table 6

The Correlation of Tests in the Battery with the Criterion of Ratings of Performance of Ward Duties

Test	Santa Clara County Hospital (N = 11)	Stanford-Lane Hospital (N = 24)
Otis	.10	.37
Strong	−.32	−.17
Super	−.32	−.25
Information	−.22	.04

Table 7

The Correlation of the Parts of the Super Test with the Criterion of Ratings of Performance of Ward Duties

	Santa Clara County Hospital (N = 11)	Stanford-Lane Hospital (N = 24)
Part I	−.47	.24
Part II	.07	−.06
Part III	−.76 ($p < .01$)	.06
Part IV	−.25	−.26
Parts I, II, III	−.38	.08
Total Test	−.322	−.25

but negative. Only the Otis is correlated positively with the criterion in both groups. However, none of the correlations are significant.

Table 7 summarizes the correlations between the parts of the Super and the criterion.

The correlations for the Santa Clara group, based on only 11 cases, are not considered as reliable as the Stanford correlations; but neither set of coefficients indicates that the Super is related positively to the criterion.

Discussion

Reliability and Validity. It has been repeatedly demonstrated that the Super technique discriminates groups; but in view of the present results, it is pertinent to question what function is being discriminated.

There are two objections to the hypotheses advanced by Super and Roper (**8**), and also by Older (**4**), that it is logical to assume the Super Test is measuring interest in the occupation itself because of its low correlations with the Strong and the Otis. First, the obtained reliability coefficient, corrected, of .70 for the total Super Test, is noticeably lower than the the coefficient of .83 previously reported (**8**). Taken together with the fact that any odd-even reliability coefficient will be too high because the assumption of uncorrelated errors is not met (**2**), the test's reliability is of such a low order as to cast serious doubt on its validity. Moreover, such low reliability would act to decrease all correlations of the Super with other tests, making the *true* relationships among them ambiguous. Since its relationship to intelligence is already significant, and that with information of near-significance, the possibility must be entertained that increased reliability for the Super-type test might show these factors to play a more significant role than has been thought.

Secondly, there is a serious danger in the reasoning of the authors cited above: if one is to ascertain what a test measures by eliminating what it does *not* measure, then all the variables which might reasonably be important should be considered. Unless this is done, any conclusion must be somewhat precarious.

Interpretation of Results. The lack of relationship of Super Test scores to each of the criteria cannot be attributed to mechanical imperfections in the assigning of grades and ratings. Space does not permit the extended treatment given this issue elsewhere (**3**); suffice it to say that inspection of the data revealed the spread of grades in the courses covered to be quite representative. Discrimination was also enhanced by the treatment of the data. With respect to the ratings, the present study is admittedly weak because of the small number of cases. The lack of a uniform rating system for all the San Jose hospitals, as well as an opportunity to instruct the raters, made the loss of so many San Jose subjects unavoidable. Although the small size and relative homogeneity of the Stanford group was regrettable, it is doubtful if many nursing schools have given as much thought and consideration to the adequacy of their rating system. Accordingly, there seems to be no reason to question the *quality* of the ratings for this group.

The Super's low reliability, with its depressant effect on the intercorrelations with the criteria, offers the most parsimonious interpretation. The amount of error in the Super is so great as to make it uncertain just what the test is measuring. Consequently, the obtained results cannot be viewed as reflecting the relationship between objective interest and achievement. The true relation will be open to investigation when two conditions are satisfied: (1) a reliable measuring instrument is made available; and (2) it is *positively* estab-

lished that the technique is measuring objective interest.

In connection with the latter point, attention should be given to the possibility that something else other than interest is being measured. As presently constituted, two-thirds of the items in the Super Test of Interest in Nursing are descriptions of pictures which may or may not have been seen in the film. All the items in Older's (4) film were of this type. Thus, the tests favor those subjects who best recognize the word descriptions corresponding to the images they have of the pictures they have seen. Super et al. make the assumption that this ability stems primarily from interest. It might prove enlightening to test the relation between a laboratory test of visual imagery and performance on the Super-type test. The relation of such performance to intelligence might be more clearly and adequately defined if the Otis were replaced by a test which measures specific aptitudes or abilities. It might also be that, in this study, the Super technique is not discriminating between more and less interested people, but, instead, between groups whose differential amounts of exposure to medical stimuli make their perception of nursing scenes proportionately more meaningful. This differential might be because of interest, but it might also be a matter of circumstance.

Implications for Future Research with the Super Technique. Future research should be concerned first with positive efforts to establish what the technique measures. The factors cited above, as well as other possibilities, should be investigated. Coordinated with this should be attempts to revise the existent Super-type tests so as to increase their reliability. In making these revisions, the results of this study indicate that multiple-choice items, used in Part I of the Nursing Test, should be given greater weight; i.e., increased in number. Not only did Part I correlate significantly with grades, but this is the most reliable type of item employed by Super. The establishment of interest as the variable being measured and the construction of a reliable instrument would then permit the investigation of the relation of objective interest to achievement.

Summary

1. The relation of scores on the Super Test of Interest in Nursing to the criteria of grades and ratings of performance of duties on the ward was not significantly above a chance level.

2. In view of the low reliability of the Super Test, the above results are ambiguous. The variable being measured may or may not be more related to the criteria. Such a determination must await a more reliable instrument.

3. It will be worthwhile to establish more positively that interest is the variable being measured. The uncertainties resulting from the technical defects of this and previous studies make desirable the further investigation of such other factors as experience and special aptitudes.

Received January 6, 1950.

References

1. Bernstein, Benjamin. *A test of interest in the metal trades.* Unpublished Master's thesis, Clark University, 1940.
2. McNemar, Q. *Psychological statistics.* New York: John Wiley and Sons, Inc., 1949.
3. Navran, L. *The relation of interest to achievement in nursing.* Unpublished Master's thesis, Stanford University, 1949.
4. Older, H. J. An objective test of vocational interest. *J. appl. Psychol.*, 1944, **28**, 99–108.
5. Roper, Sylvia A. *A test of interest in nursing.* Unpublished Master's thesis, Clark University, 1940.
6. Strip film: *Nursing as a career.* Chicago: Society for Visual Education, Inc.
7. Super, D. E., and Haddad, W. E. The effect of familiarity with an occupational field on a recognition test of vocational interest. *J. educ. Psychol.*, 1943, **34**, 103–109.
8. Super, D. E., and Roper, Sylvia A. An objective technique for testing vocational interests. *J. appl. Psychol.*, 1941, **25**, 487–498.

Acuity Differences between the Two Eyes and Job Performance

Newell C. Kephart and Joseph M. Mason

Occupational Research Center, Purdue University

The importance of vision in industry is widely recognized, and the relationships between certain visual skills and numerous aspects of job performance have been emphasized. Investigations by Kephart (**3**, **4**), Stump (**6**), and Tiffin (**7**, **8**, **9**, **10**) have convincingly demonstrated that visual skills are related to job performance, for they have found very marked relationships between these skills and hourly production, rate of labor turnover, accident-proneness, supervisors' ratings, and other acceptable industrial criteria. In general, these and numerous other studies had been concerned with individual jobs, and not with a variety of industrial jobs to determine existing basic relationships of general applicability.

To investigate these general relationships, McCormick (**5**) studied a group of approximately 5,500 employees on 92 jobs in a number of different industrial plants. From his studies he concluded that the visual acuity requirements for acceptable preformance on these 92 jobs seemed to be general and relative, and that a moderate minimum of far and near acuity in both eyes was especially pertinent to satisfactory performance. Beyond such a minimum the percentage of high criterion employees increased continuously with increase in acuity level, but at a more gradual rate. The relationship of job success to worse eye acuity was also studied, and he reports that the increase in probability of job success with increases in worse eye acuity is relatively constant for the entire range, both for near and far acuity.

McCormick, as reported, studied only the relationships between job performance and both eye acuity, near and far, and between job performance and worse eye acuity, near and far. Nowhere in the published literature on vision studies, are we able to find an analysis of the visual acuity differences between the two eyes, and the relationship of these differences to job performance. The present investigation is directed toward that objective.

Statement of the Problem

For a number of years, the Occupational Research Center (**11**) of the Division of Education and Applied Psychology, Purdue University, has been actively engaged in the development of vision-test profiles. In order to assign employees to specific jobs in which their chances of success are greatest, it is necessary to evaluate the different visual requirements of each individual job and to measure in some adequate way the visual skills of the applicant. Industrial and business organizations, subscribing to the vision service,[1] use the profiles developed in the Occupational Research Center to identify individuals whose vision does not meet the general visual demands of their jobs and who could possibly benefit from professional eye care. The profiles are also used as an aid in employee selection.

One question, unanswered by any of the previous vision studies, has to do with the relationship, if any, between job performance and acuity differences between the two eyes. In practically all vision tests, both eye acuity is measured. Then with first one eye occluded and then the other, measures of left and right eye acuities are determined. It has been observed that in a large percentage of individuals, one eye will be considerably better than the other. Is the "one-eyed" individual handicapped in an industrial job? Or does the worker with one very good eye and one very weak eye perform on a comparable level with the worker who has two very good eyes?

Thus, the present study is an extension of the research of McCormick (**5**), on the relationship between visual skills and job performance. The specific objectives of the investigation are:

1. An analysis of the over-all relationship between better eye acuity, near and far, and job performance of employees on a number of different industrial jobs.

[1] Industrial Vision Service of the Bausch and Lomb Optical Company, Rochester, New York.

2. An analysis of the over-all relationship between acuity differences in the two eyes and job performance.

3. After sub-dividing the total group into a number of sub-groups, on the basis of better eye acuity, an analysis of the acuity differences between the two eyes, as related to job performance, for each of these better eye acuity levels.

Procedures

Basic Data Used. The basic data used for this investigation were Ortho-Rater[2] test results, and job performance measures available in punched-card form in the Occupational Research Center. The data covered approximately 5,600 employees on 92 jobs in a number of different industrial establishments. With a few exceptions, they are basically the same data used by McCormick (5) for his investigation.

Vision Tests. Of the twelve vision tests included in the Ortho-Rater battery, only four were given attention in this investigation; these were: Near test (given at the optical equivalent of 13 inches), acuity, right eye, acuity, left eye; and far test (given at the optical equivalent of 26 feet), acuity, right eye, and acuity, left eye.

The better eye scores, whether left or right, for each individual, were used directly. By comparing the near acuity test scores of the left and right eyes for each individual, the one with the higher score was considered to be the better. The far acuity test scores for the left and right eyes of each individual were similarly compared and the better eye determined. Where both left and right eye had the same acuity score, the treatment was the same; for such an individual, the better eye and worse eye scores would be identical.

The difference between the better eye score and the worse eye score, for the near acuity tests, was then determined—a simple process of subtracting the one score from the other. A similar computation was made for all far acuity better-eye and worse-eye scores. The four measures of visual acuity that were actually used in the investigation were then: near acuity, better eye; near acuity, difference between better eye and worse eye; far acuity, better eye; and far acuity, difference between better eye and worse eye.

In the Ortho-Rater acuity tests, the range of possible scores is from 0 to 15. These scores can be readily converted to other measures of visual acuity, such as the Snellen notation or the AMA acuity ratings, as explained in the manual of standard practice for the Ortho-Rater (12). With the large number of test results available for this study, the distribution of better eye scores covered the entire range, although at both the lower and upper levels, the number of results included were relatively small. The greatest difference between better eye and worse eye acuities for any individual was thirteen score points.

Criteria of Job Performance. From one industrial plant to another, and from one job to another, the criteria of job performance varied considerably. In some establishments, the criteria were based on such factors as hourly production and earnings; in other plants, the measures of job performance were based upon supervisors' ratings. In general, however, and regardless of the basis of the criteria, the measurements were reduced to numerical categories, and these were subdivided into two groups, identified as the "high criterion group" employees and the "low criterion group" employees. Occasionally where there was an odd number of numerical categories, it was obviously impossible to subdivide the middle category into the "high and low criterion" groups, and it became necessary to omit these employees from consideration in the analysis. These omissions account for the apparent discrepancy in numbers of test results analyzed; from the initial group of approximately 5,600 employees, about 5,100 could be definitely assigned to one of the two criterion groups.

Data on the Relationship between Better Eye Acuity and Job Performance. In analyzing the over-all relationship between better eye acuity and job performance, summaries of the test scores and the measures of job performance were derived from the basic data. For near acuity, better eye, the per cent of all employees at a given score level who were high criterion group employees was determined. The percentages were computed for each score level for the better eye, from 0 to 15. Similar per cents were determined for each score level for far acuity, better eye. Regarding this data as only one sample out of a larger population having practically the same properties as the ones obtained, it seemed desirable to obtain "smoothed" percentages, thus minimizing irregularities in the data. This was done by a system of running averages. Guilford (2) illustrates this running-average method: for a given score, the two immediately neighboring scores are averaged with the given score, which is allowed to carry double-weight. McCormick (5), in smoothing his data, assigned equal weight to all three scores, and in this way determined the average. Both of these methods of "smoothing" were applied to a part of the data, and gave essentially the same results, so it was decided to use the latter method. The "smoothed" per cent determined for a

[2] The Ortho-Rater is a stereoscopic-type visual testing instrument manufactured by the Bausch and Lomb Optical Company, Rochester, N. Y.

given score was the quotient of the sum of the frequencies of all (both high and low criterion groups) employees for the corresponding three score levels. "Smoothed" percentages were thus determined for better eye score levels, and the resulting per cents subsequently plotted.

Data on the Over-All Relationship between Acuity Differences in the Two Eyes and Job Performance. To analyze the over-all relationship between acuity differences between the two eyes and job performance, it was necessary to derive the score difference for each individual employee from the basic data of acuity scores for the two eyes. The electric punched-card tabulation equipment available in the Occupational Research Center greatly facilitated this phase of the investigation. The approximately 5,600 cards were sorted as to better eye acuity, and then in a second sorting, according to worse eye. Then the difference between the score of the better eye and the score of the worse eye was punched on the card.

To study the over-all relationship of acuity differences between the two eyes, employees were grouped on the basis of this difference. Thus, all those showing a difference of three score points between the two eyes were placed in one group; such a category would then include an individual scoring 3 on the better eye and 0 on the worse eye, as well as an employee scoring 15 on the better eye and 12 on the worse eye, and all others where the score difference between the two eyes was three points. In similar manner, all employee test results were categorized according to the acuity differences between the two eyes. The per cent of all employees at each difference level who were high criterion group employees was determined. These percentages were computed for each acuity difference level both for the far acuity tests and the near acuity tests. To minimize the irregularities in the data due to sampling, "smoothed" percentages were computed for each difference level in the manner described above. Subsequently, these "smoothed" percentages were plotted to furnish a visual picture of the over-all relationship between acuity differences in the two eyes and job performance.

Data on the Relationship between Acuity Differences in the Two Eyes and Job Performance for Each Better Eye Acuity Level. The test scores and measures of job performance derived from the basic data were divided into 15 sub-groups on the basis of better eye acuity score. The employees represented in any given sub-group then had the same acuity score for the better eye, whereas the acuity scores for the worse eye would vary anywhere from the given acuity score of the better eye down to, and including, a score of 0. This finer breakdown into better eye acuity levels was necessary in order to study the relationship between acuity differences in the two eyes and job performance at each of these levels. For each sub-group, with a constant better eye acuity score, the per cent of the total at each difference level who were high criterion group employees was determined. Differences between the two eyes at each better eye acuity level were treated for the far tests as well as for the near tests, and the percentages of high criterion employees for each group were determined. "Smoothed" percentages were then calculated for all of these sub-groups in the manner described previously. To present a visual picture of the relationship between acuity differences in the two eyes and job performance, these "smoothed" percentages for each better eye acuity level were subsequently plotted.

Statistical Test of Homogeneity of Sample Applied to Each Sub-Group. Fisher (1) demonstrates a procedure for testing the homogeneity of a sample in a $2 \times n'$ classification which is applicable to the data used in this investigation. This test of homogeneity was applied to all sub-groups, both for the near and far tests, where the total N was sufficiently large to make the test applicable. This resulted in an elimination of only those sub-groups where the better eye acuity score was 3 or less. For all other sub-groups where the total for any given difference level was less than 15, the values were combined with the succeeding difference groups to bring the total to the required 15. In determining the proportions, calculations were carried to seven decimal places, as Fisher states (1: 91) that using five decimal places the value of χ^2 is not quite correct in the second decimal, and to avoid doubts as to the precision of calculation, two more places are desirable.

All χ^2 values determined after applying this test of homogeneity were interpreted in terms of P from the values given in Fisher's Tables (1: 104–105), and subsequently each χ^2 value was compared to the .05 probability values as given in the Tables. This comparison was made because Fisher states (1: 82): "We shall not often be astray if we draw a conventional line at .05, and consider that higher values of χ^2 indicate a real discrepancy."

Results

Relationship between Better Eye Acuity and Job Performance. The relationship existing between better eye acuity and job performance is reflected by the percentage of individuals at each score level who were in the high criterion group of employees. The basic relationship represented by these percentages is shown

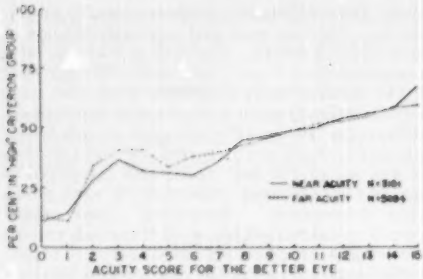


FIG. 1. Relationship between better eye acuity and job performance.

graphically in Figure 1, which presents the relationship for all of the 92 jobs combined, both at the near and far acuity distances.

Both the far and near acuity, better eye, trend lines show a relatively marked increase in the percentage of high criterion group employees at each successive better eye score value. Examination of Figure 1 shows that for all 92 jobs combined there is a general increase in the proportion of individuals who were high criterion group employees at successively higher scores, and that the increase is quite persistent throughout the entire range of scores, though perhaps a little more marked in the lower score range. The few irregularities in the general trend line are probably due to sampling errors, for even though the total N is large, the number of employees scoring in the 0 to 3 and 13 to 15 score levels are relatively small. The general trend presented here graphically would indicate that at least a moderate degree of visual acuity in one of the two eyes, both at near and far distance, is necessary for satisfactory performance on the types of industrial jobs covered by this investigation. Further, it is indicated that the greater the degree of visual acuity in the better eye the greater the probability of success for the employee on these jobs.

The general relationship existing between the better eye acuity and job performance, as indicated in Figure 1, is similar to the relationship determined between both eye acuity and job performance (**5**). This is true both for the near and far acuity distances.

Over-All Relationship between Acuity Differences in the Two Eyes and Job Performance. The relationship existing between the visual acuity differences in the two eyes and job performance is reflected by the percentage of employees at each difference level who were in the high criterion group of workers. The basic relationship represented by these percentages is shown graphically in Figure 2. In this figure is presented the relationship for all of the 92 jobs combined, both at the near and far acuity distances.

The acuity differences represented cover the entire range of acuity differences between the better eye and the worse eye. Thus, at any given score difference on the base line, the value plotted directly above it represents the percentage of the total group having that acuity difference who were in the high criterion group of employees. For example, above the point labeled "7" on the base line, are represented all employees showing a difference of 7 score points between the better eye acuity score and the worse eye acuity score; it would include those who scored 7 on the better eye and 0 on the worse eye, as well as those who scored 14 on the better eye and 7 on the worse eye, and all others whose acuity difference between the better and worse eyes was 7 score points.

For acuity differences between the two eyes, there appears to be no relationship between these differences and job performance, indicating that with greater differences between the two eyes (both near and far), the probabilities of satisfactory performance on industrial jobs of the types investigated here are approximate-

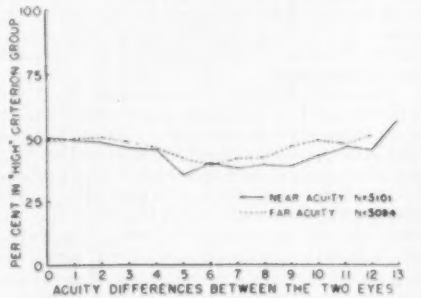


FIG. 2. Over-all relationship between acuity differences in the two eyes and job performance.

ly the same as the probabilities of satisfactory performance where the two eyes are equally good. The general trend presented here graphically indicates the over-all relationship for acuity differences between the two eyes and job performance, and suggests that such acuity differences do not, at any difference level, significantly affect the probabilities of satisfactory performance for the combined group of jobs.

Relationship between Acuity Differences in the Two Eyes and Job Performance for Each Better Eye Acuity Level. For a given better eye acuity, the relationship existing between acuity differences in the two eyes and job performance is reflected by the percentage of employees at each difference level who were in the high criterion employee group. The basic relationships are presented graphically for better eye acuity scores 3 to 15 in Figures 3 to 6, inclusive.[3] These figures represent the relationships for all of the 92 jobs combined, both at the near and far acuity distances.

At any given better eye acuity, the apparent lack of relationship from the low acuity differences to the high acuity differences indicates that such differences do not significantly affect the probabilities of satisfactory performance for these jobs. Where the acuity difference between the two eyes is large, the probability of satisfactory performance on these jobs is approximately the same as the probability of satisfactory performance where the difference is small. This general relationship holds for all better eye acuity levels, both near and far.

Results of Statistical Test of Homogeneity of Samples. Fisher's test of homogeneity of a sample in a $2 \times n'$ classification, applied to each sub-group for each better eye acuity score, substantiates the basic relationships indicated in the graphic presentations. None of the proportions of high criterion employees within a given better eye acuity sub-group deviates significantly from the theoretical proportion for the group as a whole. The χ^2 values obtained, the P values interpolated from Fisher's Tables (1:104–105), and the χ^2 values for a P of .05 are presented in Tables 1 and 2,[4] for far and near acuity distances. Fisher states that if P is between .1 and .9 there is certainly no reason to suspect the hypothesis tested, or that the observed deviations could not have occurred by chance. We have therefore, accepted his suggestion and hold that only where P is smaller than .05, is there sufficient indication of a significant deviation. None of the P values for any better eye acuity level, near or far, is as small as .05; the range for the 24 sub-groups is from .09 to .82, with 23 of the 24 falling within the .1 to .9 limits mentioned above.

This statistical test then confirms the observed relationship between acuity differences of the two eyes and job performance, at constant better eye acuity levels. For the 92 industrial jobs combined, as studied in this investigation, acuity differences are not related to probabilities of satisfactory job performance.

Summary

An investigation was made of the existing relationship between certain visual skills and job performance on a variety of industrial jobs. Ortho-Rater scores of approximately 5,600 employees on 92 different industrial jobs comprised the basic data. Using various measures of job performance as criteria, about 5,100 of these employees were identified as either high criterion group employees or low criterion group employees. Four measures of visual acuity were given particular attention and their relationship to job performance analyzed. These four acuity measures were: better eye acuity, near; better eye acuity, far; acuity differences between the two eyes, near; and acuity differences between the two eyes, far. Acuity differences between the two eyes were analyzed both for an over-all relationship to job performance, and for each better eye acuity level.

The following conclusions are drawn from this investigation:

1. For jobs of the type studied, a moderate minimum of visual acuity in one eye or the

[3] To reduce printing costs, Figures 3, 4, 5, and 6, and Tables 1 and 2 have been deposited with the American Documentation Institute and may be ordered as Document 2870 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $0.50 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $0.60 for photocopies (6 × 8 inches) readable without optical aid.

[4] See footnote 3.

other is necessary for satisfactory job performance.

2. Beyond this minimum, probability of satisfactory performance increases quite constantly with greater degrees of acuity in the better eye.

3. In the over-all relationship between acuity differences in the two eyes and job performance, such differences are not significantly related to job performance.

4. A horizontal straight-line relationship exists between acuity differences in the two eyes and job performance, so that the probability of satisfactory performance where acuity differences are large is approximately equal to the probability of satisfactory performance where acuity differences are negligible.

5. For any given better eye acuity level, acuity differences in the two eyes do not affect the level of job performance for the types of jobs studied.

This investigation has been primarily concerned with the over-all relationships for the 92 jobs combined. It is possible that on certain industrial jobs that require particular types of visual skills, acuity differences in the two eyes may be related to job performance. An extension of this investigation to employees on particular jobs would be necessary to determine these relationships.

Received January 23, 1950.

References

1. Fisher, R. A. *Statistical methods for research workers.* (4th Ed.) London: Oliver and Boyd, 1932.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill Book Company, Inc., 1942.
3. Kephart, N. C. An analysis of professional eye care and industrial efficiency. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1946, **51**, 166–170.
4. Kephart, N. C. Visual skills and labor turnover. *J. appl. Psychol.*, 1948, **32**, 51–55.
5. McCormick, E. J. *An analysis of visual requirements in industry.* Ph.D. Thesis, Purdue University, 1948.
6. Stump, N. F. Vision tests predict worker capability. *Factory Mgmt. and Maint.*, 1946, **104**, 121–124.
7. Tiffin, J. *Industrial psychology.* (2nd Ed.) New York: Prentice-Hall, Inc., 1947.
8. Tiffin, J. The use of visual data as an aid to increase production and efficiency. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1944, **49**, 1–4.
9. Tiffin, J., and Greenly, R. J. Employee selection tests for electrical fixture assemblers and radio assemblers. *J. appl. Psychol.*, 1939, **23**, 240–263.
10. Tiffin, J., and Wirt, S. E. The importance of visual skills for adequate job performance. *J. consult. Psychol.*, 1944, **8**, 80–89.
11. Wirt, S. E. Statistical laboratory for vision tests at Purdue University. *J. appl. Psychol.*, 1946, **30**, 354–358.
12. *Standard practice in the administration of the Bausch and Lomb occupational vision tests with the Ortho-Rater.* Bausch and Lomb Optical Company, Rochester, N. Y., 1944.

Point Centering of Signals on an Area *

Adelbert Ford, David Rigler, and Genevieve E. Dugan

Lehigh University

This task relates to the precision with which signals, consisting of white spots or patches, can be located or "tracked" on a dark field, such as that of a radar scope, by means of a device by which an operator can move a point cursor to the center of the signal, using a pantograph lever system giving two degrees of muscular freedon, and experimental amounts of movement reduction.

The problem includes the discovery of the absolute amount of the error, its comparison with the sizes of errors previously reported for scale reading techniques of locating signals, the presentation of any tendencies towards drifts and other variability in efficiency, and some possible causes for the trends.

The necessity of precise signal location on scope faces is a problem of human operation which is critical in the technical methods of obtaining navigational information for the use of ships and commercial aircraft traveling in congested regions at night and in weather conditions adverse to good visibility.

History of the Problems

A signal can be located on a scope face by referring it to a system of scaling lines. The configurational and systematic errors of this procedure have already been reported by Ford (**5**, **6**) and by Reed and Bartlett (**9**). The average random error for this method has been shown to be about 0.020 inch of scope distance for clear signals, and about 0.040 inch for typical unclear signals. The systematic errors are too great to be tolerated. Individual differences are relatively small.

Signals can also be located by a superimposed cursor which actuates a remotely controlled indicator. In this situation there are two human error problems: (a) how accurately the operator sets the cursor on the signal of "pip," and (b) how accurately he reads the indicator dial. The first of these problems will be partially answered in the present article. With regard to the second, Chapanis (**2**, **3**) and Long (**7**) have shown that a direct-reading counter is better than a scale with an indicator needle when reading the numerical value of the position of a signal which is already established as to location, but not for the purpose of moving an indicator to a specified numerical position.

Most of the war instruments used cursors with hair lines for centering the x- and the y-axes by separately controlled hand wheels. The use of a pantograph lever system for moving a pointer cursor with two degrees of muscular freedom was adopted by Reed and Bartlett (**8**) on PPI traffic survey scopes, and although they did not have a precision method for registering errors, they reported the centering procedure to be about twice as precise as the scale-reading method. This report is in the direction of our own findings on the sector scopes. In using cursors on changing target positions Ellson and Wheeler (**4**) reported the "range effect," which is a tendency of the operator to "overshoot" when the movement of a target position is less than expected, and to "undershoot" when the movement is greater than expected. Bartlett and Sweet (**1**) noted that when two operators were simultaneously tracking the same target their agreement in readings became poorer and poorer as the successive target positions moved more rapidly.

Apparatus

Figure 1 is a reproduction of an artificial signal (the oval area in the center) surrounded by eight station positions, actual size. In some of the experiments a photographic reproduction of a real radar "pip" was substituted for this artificial signal, at approximately the same size of signal area.

The cursor was a black dot the size of a No. 60 B and S drill inscribed on the under side of a transparent plastic plate. This cursor plate could be moved from any specified station position to the center of the signal by means of a pantograph system of levers such that experi-

* This research was executed under Contract No. W33-038-ac-22561 between the Institute of Research, Lehigh University, and the USAF Air Materiel Command, Aero Medical Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.

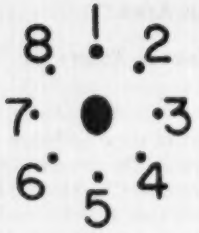


FIG. 1. Artificial signal, the oval area in the center. Numbered dots are the station positions from which the operator moves the point cursor (actual size).

mental ratios of knob movement to cursor movement could be established at values of 1:1, 2:1, 4:1 or 6:1.[1] The vertical error of the centering was indicated by an attached rotating mirror which moved a light beam along a horizontal scale at a magnification of 100, and measured variations in settings to a precision of 0.001 inch.

The signal was covered by a solenoid-actuated shutter. The subject could open the shutter at his own discretion, but the shutter would close after an experimentally determined number of seconds allowed for centering of the target.

[1] For an expanded description of the apparatus, and additional tables and charts of results, see Ford, A., Rigler, D., and Dugan, G. E., Pantograph Radar Tracking: Point Centering Experiments, USAF Technical Report No. 5969, Air Materiel Command, Aero Medical Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.

The target board was mounted thirty degrees from horizontal and illuminated with thirty foot candles of well diffused tungsten and fluorescent light.

Experimental Procedure

Method of Average Error. In this series the subject made his own settings of the cursor on the target. The order of the various conditions was determined by the Latin Square arrangement to equate fatigue and training effects from one condition to another.

The statistical treatment is determined by the fact that we discovered that all subjects tended to show "drifts" up and down for what they considered the center of the target area. A method of running averages and running standard deviations was adopted.

Table 1

Point Centering Experiments, Artificial Pip: Subject WEM

(Values expressed in decimal fractions of one inch)

Kind of Tracker	Ratio	Time Limit in Seconds	Stand. Dev. Total	Ave. Running S.D.	Max. Dev. Running Average	Drift Index
Pantograph	6/1	2	.0038	.0026	.0088	3.4
Pantograph	4/1	2	.0026	.0020	.0045	2.3
Pantograph	2/1	2	.0027	.0024	.0029	1.2
Fing. Cent.	1/1	8	.0028	.0020	.0065	3.3
Fing. Cent.	1/1	6	.0031	.0024	.0053	2.2
Fing. Cent.	1/1	4	.0031	.0025	.0049	2.0
Fing. Cent.	1/1	2	.0036	.0028	.0028	1.0

Discrimination Limen (Weber's Law), S.D. = .0010
S.D., Reading from Scale A = .0166.

Twenty-four trials were allowed for practice. Then about 180 settings were made in a series under a given experimental condition. The average and the standard deviation was computed for the first 24 readings. Then the first eight were dropped, and the next eight readings added, and the next average and standard deviation noted. This was done successively until the last 24 readings out of the 180 were exhausted. Table 1 and Figure 2 have been designed on this kind of data. Then the average and the standard deviation of the entire series was computed.

The drift phenomenon required quantification, so we established a "drift index" which was the maximum change in the running average divided by the average running standard deviation. If the operator's running average

shifted from a −0.003 to a +0.003 inch, and his average running standard deviation was 0.002 inch error, his drift index would be 3.0. This example is fairly close to what actually happened. (See Figure 2.)

The Method of Constant Stimuli. Where the subject makes his own attempt to place the cursor in the center of a signal area one may consider three causes for the commission of an error: (a) the subject hasn't the motor skill to make a precise positioning of the cursor point, (b) the subject hasn't the fineness of visual perceptual discrimination to see where the center is, or (c) the subject lacks an attitude of working to precision and works only to the standard he thinks will satisfy the supervisor.

To get evidence on the second of the above factors, the experimenter set the cursor point in chance order above and below the actual center, and the subject reported "higher" or "lower" in a typical Weber-Fechner technique. The experimenter started with a plus-or-minus 0.004, then 0.003, and so on down to 0.001 inch differences. For this purpose the difference limen was considered to be that value which the subject reported correctly by an amount equal to the standard deviation of the probability curve (68 per cent instead of the usual 50 per cent). The value of this limen, for WEM, is given at the bottom of Table 1. Such a value for each subject was assumed to yield a measure of the pure perceptual factor uncontaminated by the factor of motor skill.

Scale Reading Comparison. Using a set of photographic reproductions of artificial pips of the same dimension as that shown in Figure 1, superimposed on the multiple scaling system like that used in previous experiments which yielded the highest accuracy, each subject was required to estimate elevation of the signal above or below a line of zero reference, by the use of scaling lines only. Ninety-six readings were taken for each subject. The value is indicated under Table 1 after "S. D., readings from Scale A." This was done to show the difference in precision between scale-reading method and cursor-setting method, with individual differences held constant.

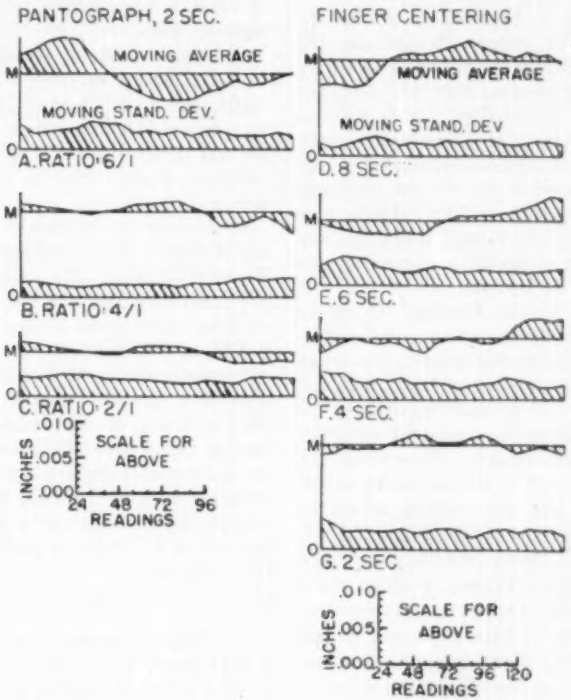


FIG. 2. Chart of error trends for Subject WEM on artificial signal.

Results

1. *Subjective Drift.* The tendency to wobble back and forth, above and below, the mean center, with the majority of readings too high for a while, then too low for a while, is pictured in Figure 1, for WEM. All seven subjects did this to some degree on both the artificial signals and the reproduction of the natural signal. The drift index was as much as *three* for every subject on one or the other of the runs. We are inclined to believe that this tendency to wobble around a center is characteristic of nearly any difference limen experiment by the method of average error, but has been heretofore hidden by the use of gross averages and gross measures of dispersion. About 0.003 inch wobble is the expectancy for this size of signal.

2. *Pantograph Reduction Ratio.* The best reduction ratio for hand movement in relation to cursor movement was close to 2:1. This is shown pictorially in Figure 2 for WEM, and in Table 1. The criteria for making this statement are: (a) least amount of drift, (b) smallest average running standard deviation of errors, and (c) smallest total standard deviation for errors. On the artificial signals, two out of the five subjects did almost equally well on the 4:1 lever ratio. On the reproductions of natural signals all three subjects were superior on the 2:1 ratio. Since the last situation is closer to the job requirements, the vote must go to the 2:1 ratio.

3. *Comparison to Scale Reading.* Using the total standard deviation (which includes variability due to drift) as the criterion for cursor centering by the 2:1 pantograph ratio, all five subjects showed a heavy superiority for pantograph centering as compared with the technique of scale reading. Expressing the standard deviation of cursor centering errors as percentage of the scale reading errors, we obtained the following values: 16, 46, 35, 44, 27. Over all, this means that the expectancy is that pantograph centering is about three times as accurate as reading the same signal on a superimposed scale (with much less mental effort). If a method of reducing drift could be found, the advantage would be still greater.

4. *Hand Positioning.* Allowing the subject to rest the palm of his hand on the table while making a final ecntering resulted in a slightly superior score, statistically, but the difference was very small and of doubtful statistical significance. Subjects became rapidly antagonistic to the compulsory free arm method, without the palm rest, however, and reported that it was a strain to do it in this manner.

5. *Preceptive versus Motor Causes of Error.* The limen of perceptive discrimination, as indicated by the method of constant stimuli, is only a half to a third the value of the total standard deviation of cursor centering errors. This means that perceptive errors account for only a small part of the total error by the pantograph centering technique. We may probably make a large allowance for motor skills and precision attitude.

6. *Effect of Time Restriction.* A point cursor can be centered on a signal in a time as short as two seconds before any appreciable increase in error is generated. This was true of all eight subjects.

7. *Type of Cursor.* When the point cursor was surrounded by rings a tenth of an inch apart, like a "bull's eye," the drift tendency was reduced for all three subjects on the natural signal series. This suggests that the problem of drift can be solved by cursor design.

8. *Absolute Size of Errors.* The total standard deviation of pantograph centering errors for the 2:1 lever ratio (which includes the drift tendency) for the five subjects on the artificial signals was a plus-or-minus 0.0027, 0.0047, 0.0050, 0.0063, and 0.0026 inch, or an average expectancy of four thousandths of an inch for a signal of this size. The values for the centering of the reproductions of the natural pips were 0.0068, 0.0070, and 0,0050, or an average expectancy of six thousandths. Previous work by Ford (**5**, **6**) has shown that natural signals are about twice as difficult to localize on a surface as artificial signals with sharp contours.

Summary

1. Centering signals on an area by means of a pantograph-controlled cursor is from two to four times as precise as the location of such signals by scale reading methods.

2. Using a point cursor, the subjects tend to exhibit systematic drift back and forth across the signal center.

3. A reduction ratio of two units of hand movement to one unit of cursor movement is near the optimum.

4. The use of a palm rest resulted in a slight superiority of scores in centering, and was greatly preferred by all subjects.

5. Centering in two dimensions can be done in a minimum of two seconds before reaching definite increases in error.

6. The limits of visual perception account for less than half of the range of errors. Skill and precision attitude may account for the remainder.

7. A point cursor surrounded by "bull's eye" rings reduces the tendency to drift around actual center.

8. With a signal 0.261 inch in length, the standard expectancy in centering is a plus-or-minus error of 4 thousandths of an inch for artificial signals, and six thousandths of an inch for reproductions of real radar pips. Individual differences are relatively large.

Received February 9, 1950.

References

1. Bartlett, N. R., and Sweet, A. L. *The agreement in range and bearing reports from two VF Radar Repeaters tracking the same target simultaneously with the B Scope.* Memorandum Report No. 166-I-12, Special Devices Center, U. S. Navy, 1 November 1946, restricted. Johns Hopkins University.
2. Chapanis, A. *Speed of reading target information from a direct reading counter type indicator versus conventional radar bearing and range dials.* Memorandum Report No. 166-I-3, Special Devices Center, U. S. Navy, 1 November 1946, restricted. Johns Hopkins University.
3. Chapanis, A. *The relative efficiency of a bearing counter and bearing dial for use with PPI Presentations.* Memorandum Report No. 166-I-26, Special Devices Center, U. S. Navy, 1 August 1947, restricted. Johns Hopkins University.
4. Ellson, D. G., and Wheeler, L. *The range effect.* Technical Report No. 5813, May 1949, Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio. University of Indiana.
5. Ford, A. Types of errors in location judgments on scaled surfaces. I. Errors of configuration. *J. appl. Psychol.*, 1949, **33**, 373–381.
6. Ford, A. Types of errors in location judgments on scaled surfaces. II. Random and systematic errors. *J. appl. Psychol.*, 1949, **33**, 382–394.
7. Long, G. E. *Speed and accuracy of readings as a function of design in the sensitive airspeed indicator* Technical Report No. 5836, August 1949, Air Materiel Command, Wright-Patterson Air Force Base, Dayton, Ohio.
8. Reed, J. D., and Bartlett, N. R. *Comparison of manual and standard methods of target identification.* Memorandum Report No. 166-I-9, Special Devices Center, U. S. Navy, 1 February 1947, restricted. Johns Hopkins University.
9. Reed, J. D., and Bartlett, N. R. *The accuracy of range information on a PPI with respect to the distance of the target from the range rings.* Memorandum Report No. 166-I-29, Special Devices Center, U. S. Navy, 1 November 1947, restricted. Johns Hopkins University.

Influence of Friction in Making Settings on a Linear Scale *

William Leroy Jenkins, Louis O. Maas, and David Rigler

Lehigh University

In a previous study of the process of making settings on a linear scale by means of a control knob (1), a number of variables were investigated. The most significant of these turned out to be the *ratio* between pointer-movement and knob-turn. A ratio providing one or two inches of pointer-movement for one complete turn of the knob appeared to be optimal. Finer ratios wasted time and effort in traveling to the approximate location. Coarser ratios were poorly adapted to making the final adjustment.

In the apparatus employed for these studies there is a noticeable resistance with extremely coarse ratios. The question arose: Would the optimal ratio be changed by artifically equalizing the friction at all ratios? The present study is concerned with this question, and also with the more general problem of the influence of added friction at the optimal ratio.

Apparatus and Procedure

The apparatus and procedure were essentially the same as described in the previous study (1). The subject *matches* the position of a lighted insert in a black bakelite scale with a pointer controlled by a rotary knob. The permitted error-tolerance is determined by the width of the pointer in relation to the width of the lighted insert. In the present investigation, the permitted error tolerance was .007″; (insert .032″; pointer .025″), and the knob diameter 2¾″.

By means of two chronoscopes, time is measured separately for travel to the approximate location and for making the final adjustment. Similarly, action potentials from the active forearm are accumulated and measured separately during travel and during final adjustment. Mean travel time is computed for two standard distances: 10 sixteenths and 50 sixteenths of an inch. Mean total time is then computed as mean travel time plus mean adjusting time. Similarly computations are made for action potentials, which are recorded in meter-scale readings having no intrinsic significance but comparable for the same subject.

In the present experiments, friction was added by means of a Prony brake on the shaft immediately behind the control knob. For the series on equalized friction, the braking was adjusted at each ratio so that a pull of 300 grams was required at the periphery of the 2¾″ knob—the same as for the coarsest ratio used. For the series on the influence of friction at the optimal ratio, sufficient braking was applied to require pulls of 100, 400, 700, 1,000, and 1,300 grams at the periphery of the knob.

Four subjects were used, all young men students. One of them (DMS) had served in several previous experiments. The other three were new.

Results—Equalized Friction

Table 1 shows mean total time and mean total potential at six different ratios without equalizing friction. These results are analogous to those reported in the earlier study (1). The optimal ratio is in the region of one or two inches of pointer movement for one complete turn of the knob. With a travel distance of 10 sixteenths of an inch, the ratio of 1.18 is optimal. When the travel distance is 50 sixteenths of an inch, the ratio of 2.42 is slightly better.

Table 2 gives mean total time and mean total potential for the same six ratios with friction artificially equalized. At 10 sixteenths travel distance, there is now little to choose between ratios 1.18 and 2.42. At 50 sixteenths travel distance, the ratio 2.42 becomes significantly better.

The two sets of data shown in Table 1 and Table 2 were taken at different times; so the absolute values cannot be directly compared.

* This research was executed under Contract No. W33-038-ac-22561 between the Institute of Research, Lehigh University, and the USAF Air Materiel Command, Aero Medical Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio.

Table 1

Influence of Ratio on Time and Potential—Friction Not Equalized

	Mean Total Time							
	10 Sixteenths Travel				50 Sixteenths Travel			
Ratio	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
1.18	*16.3*	*26.8*	*24.8*	*26.7*	*24.3*	34.0	30.4	*33.5*
2.42	19.1*	27.5	25.3	32.6*	27.1*	*31.5*	*29.3*	36.6
4.08	19.2*	30.9*	26.4	35.2*	23.6	34.1	29.6	38.8*
6.28	19.5*	33.8*	28.8*	33.9*	23.5	35.8*	32.0	36.7*
9.70	23.8*	34.8*	31.6*	45.2*	26.6	37.2*	34.8*	48.4*
16.3	32.8*	48.8*	37.7*	52.5*	34.4*	50.4*	39.7*	56.1*
	Mean Total Potential							
	10 Sixteenths Travel				50 Sixteenths Travel			
Ratio	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
1.18	*14.4*	*18.1*	19.7	21.5	*23.2*	25.7*	28.5*	34.7*
2.42	17.1*	18.1	*18.1*	*21.4*	25.1	*21.7*	*22.9*	*27.8*
4.08	16.5*	21.0*	19.6	23.7	21.3	24.2*	23.6	29.7
6.28	18.1*	24.9*	23.6*	26.0*	22.1	27.7*	28.4*	30.0
9.70	19.7*	23.8*	27.0*	36.6*	23.3	26.2*	31.8*	41.0*
16.3	24.9*	38.1*	41.1*	45.1*	26.9*	40.9*	44.3*	49.5*

Tolerance: .007". Knob diam.: 2¾".
* Significant difference beyond 1% level from italicized figures.

Table 2

Influence of Ratio on Time and Potential Friction—Equalized at All Ratios

	Mean Total Time							
	10 Sixteenths Travel				50 Sixteenths Travel			
Ratio	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
1.18	26.8	27.8	*20.2*	*40.4*	41.2*	44.2*	28.6	55.2*
2.42	*24.9*	*26.6*	21.3	41.8	*31.3*	*34.6*	*26.5*	*48.6*
4.08	28.0	29.8	21.9	43.4	34.4	34.6	25.5	49.4
6.28	30.3*	30.8*	24.7*	47.5*	35.4*	35.2	27.7	52.3
9.70	31.9*	35.1*	27.4*	54.9*	36.5*	38.3	30.0*	59.3*
16.3	44.8*	38.6*	35.8*	70.9*	47.4*	41.8*	37.8*	73.3*
	Mean Total Potential							
	10 Sixteenths Travel				50 Sixteenths Travel			
Ratio	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
1.18	24.6	*19.2*	19.9	31.0	42.2*	36.4*	33.5*	46.0*
2.42	*23.0*	19.7	*19.2*	*27.8*	*33.0*	*29.1*	*28.0*	*37.4*
4.08	25.2	22.3*	19.7	32.1	32.4	27.9	25.7	39.3
6.28	28.3*	25.3*	21.9	33.9*	33.9	29.3	27.1	39.1
9.70	31.2*	33.8*	24.1*	41.1*	35.3	37.0*	27.7	46.3*
16.3	45.2*	29.9*	31.4*	59.3*	48.8*	33.1	34.6*	61.7*

Tolerance: .007". Knob diam.: 2¾".
* Significant difference beyond 1% level from italicized figures.

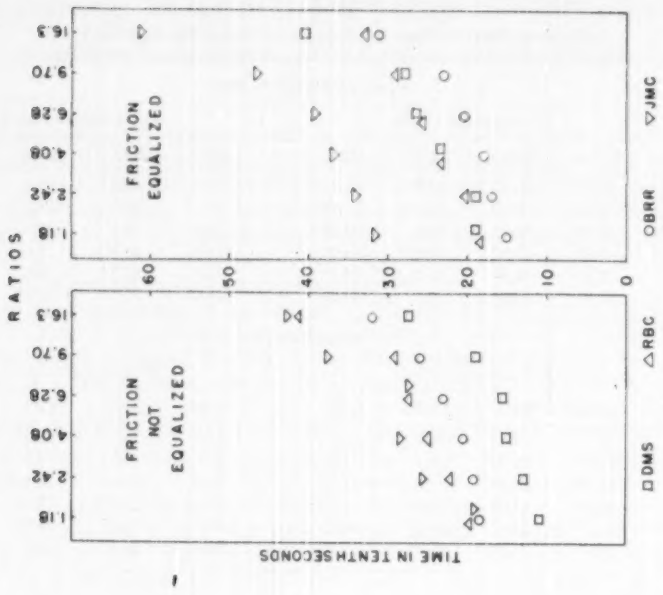


FIG. 2. Influence of ratio on adjust time.

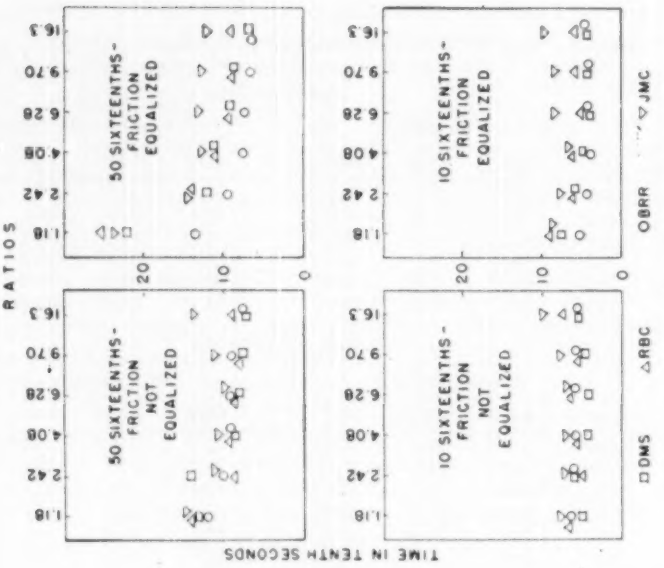


FIG. 1. Influence of ratio on travel time.

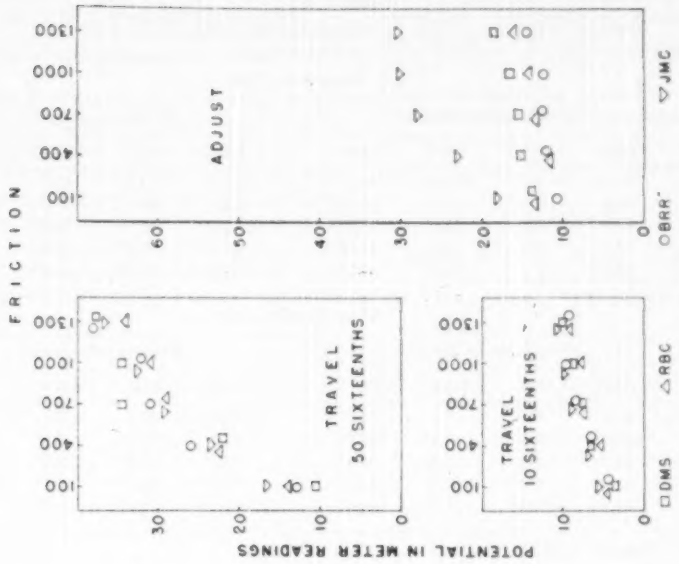


FIG. 4. Influence of friction on potential.

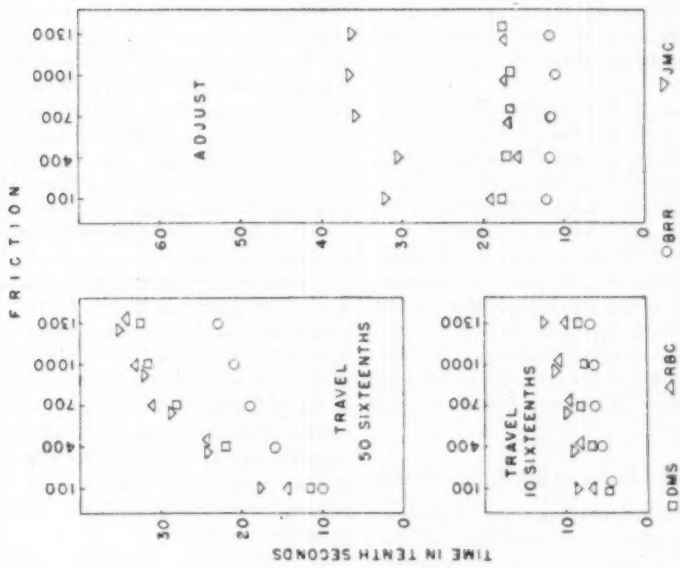


FIG. 3. Influence of friction on time.

Table 3

Influence of Friction on Time and Potential at the Optimal Ratio (1.18)

	Mean Total Time							
	10 Sixteenths Travel				50 Sixteenths Travel			
Friction	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
100	*22.4*	*26.1*	*16.6*	*40.3*	*29.2*	*33.7*	*22.2*	*49.5*
400	23.8	23.9*	17.4	39.1	39.0*	40.3*	27.8*	54.3*
700	24.5	26.3	18.0	45.4*	44.5*	47.9*	30.8*	64.2*
1,000	23.7	28.0	17.6	47.6*	48.1*	50.2*	32.0*	68.4*
1,300	25.8*	27.6	18.8*	48.7*	51.0*	51.6*	34.8*	71.1*
	Mean Total Potential							
	10 Sixteenths Travel				50 Sixteenths Travel			
Friction	DMS	RBC	BRR	JMC	DMS	RBC	BRR	JMC
100	*17.5*	*17.9*	*15.1*	*23.7*	*24.3*	*27.5*	*23.9*	*34.5*
400	22.0*	17.0	18.9*	30.9*	37.2*	34.2*	38.1*	46.5*
700	23.0*	20.7*	21.4*	36.9*	49.8*	42.3*	43.8*	56.9*
1,000	25.1*	22.0*	21.9*	39.7*	55.1*	44.2*	44.3*	62.5*
1,300	28.5*	25.3*	23.5*	40.6*	56.1*	50.1*	52.3*	66.5*

Ratio: 1.18. Tolerance: .007". Knob diam.: 2¾".

* Significant difference beyond 1% level from italicized figures.

However, it is evident that equalizing the friction at all ratios does not change the region of the optimum at short travel distances and has only a moderate effect at longer travel distances.

Figures 1 and 2 show graphically why this is true. In Figure 1, travel time is compared with and without equalized friction. Only in the drop from 1.18 to 2.42 at 50 sixteenths travel is there any marked difference. In Figure 2, adjusting time is compared in a similar manner. With or without equalized friction, the ratio of 1.18 gives the shortest adjusting time. Any advantage of a higher ratio is, therefore, traceable to the slightly faster travel time over long distances. In most ordinary adjustments, the ratio of 1.18 remains optimal regardless of whether friction is artificially equalized.

Results—Added Friction at the Optimal Ratio

Normally in the apparatus a pull of approximately 100 grams at the periphery of the 2¾" knob is required to move the pointer. By means of the Prony brake, friction was added so that pulls of 400, 700, 1,000, and 1,300 grams were required.

Mean total time and mean total potential are shown in Table 3. At the shorter travel distance, even the highest level of friction increases the total time only slightly in three of the four subjects. The effect on total potential is more marked. At the longer travel distance, the slowing effect of friction is noticeable even at 400 grams and becomes progressively greater. Total potential increases in an analogous fashion.

Figures 3 and 4 show why these phenomena occur. In Figure 3 it will be noticed that added friction has its main effect on travel time, while adjusting time is scarcely affected at all. In fact, adjusting time with friction at the 400 gram level is slightly, though not significantly, shorter than at 100 grams. In Figure 4, travel potential shows a noticeable increase with added friction, but adjusting potential rises sharply in the case of only one of the four subjects.

In their reports, subjects do not complain about added friction at 400 grams but find

the higher level distinctly irksome even when no marked slowing of settings is evident.

Summary

Previously it had been shown that a ratio of one or two inches of pointer movement for one complete turn of a control knob was optimal for matching settings on a linear scale. In the present study, friction was artificially equalized at all ratios by means of a Prony brake. Even with equalized friction, the optimal ratio was unchanged for short travel distances. With large amounts of travel before setting, the optimal ratio was slightly increased. The influence of added friction at the optimal ratio was also studied. Added friction, even in excessive amounts, has no effect on adjusting time but does increase travel time and the action potentials from the active forearm.

Received February 9, 1950.

Reference

1. Jenkins, W. L., and Connor, M. B. Some design factors in making settings on a linear scale. *J. appl. Psychol.*, 1949, **33**, 395–409.

The Galvanic Skin Response as a Test of Advertising Impact *

Edwin Golin and Samuel B. Lyerly

The University of North Carolina

In preparing an advertising campaign, the advertising agency is usually faced with the job of selecting the most effective layout for publication. The copy department submits several layouts concerning the product to be advertised and from this group, one arrangement with the greatest potential "impact" upon prospective consumers must be chosen. A misjudgment in this selective process may involve a large loss of capital, time, and energy. To avoid this waste, many techniques have been employed in an attempt to make the selection less haphazard. Some of the methods used are subjective judgments of large samples of population, experts' judgments, coupon returns, and so on. However, none of these methods has proved completely satisfactory because of either low validity or the relatively great expense and time factors involved. Experiences of this sort have placed the advertising industry on the alert for new methods of testing the impact of advertisements before the publication stage.

One possible way of approaching this problem is through the use of the galvanic skin response (GSR). Briefly, the GSR results from changes in skin resistance to an induced current. These changes of resistance may be read off a scale by deflections from a mirror galvanometer. It is assumed by many investigators that the GSR is a reflection of general autonomic nervous activity and that autonomic activity accompanies emotional states; hence GSR may be considered as one possible index of emotion. Applied to an advertisement, then, a layout of highly affective components for any particular observer should result in greater GSR deflections than a less affective one. The GSR method, if found valid, has the advantage of being a low-cost, time-saving technique.

With this in view, investigators have begun experimenting with various types of GSR apparatus and advertising media. Among the limited amount of research in this area, the most prominent is the study by Eckstrand and Gilliland (1). These investigators found a significant agreement between GSR changes to a series of advertisements and the sales effectiveness of those advertisements. They concluded that the psychogalvanic method of testing advertising material could be a predictive technique if handled under properly controlled conditions.

It was because of this rapidly growing interest in GSR as a copy testing method that the authors in conjunction with an advertising corporation attempted to set up a preliminary factorial experiment that would indicate whether the GSR of a group of subjects would show significant differences when a series of advertisements concerning the same product was presented. This was to be a basic study and, as such, was not to involve other criteria such as subjective reports, sales tests, coupon returns, and so on. If the GSR did indicate significant differences among advertisements, then this could be used as the basis for further investigation and interpretation.

The Experiment

Advertising Material. Three products were chosen (nose tissue, ice cream, and air travel) which were considered within the range of interests of the subjects tested. These products were given fictitious brand names. Four advertisements were prepared for each product, varying the arrangement, or layout. Each ad consisted of a short slogan (three to six words), a picture, and the "brand" name. Layout A, for each product, featured prominently a picture of the article or service involved (for example, a large box of nose tissue or a picture of an airliner). Layout B included a picture of a single person using the advertised product (e.g., a boy eating an ice cream cone). Layout C portrayed two human figures, one male and one female, in poses which were

* The authors wish to thank N. W. Ayer & Son, Inc., Philadelphia, for financial aid which made this study possible. Special thanks are due Mr. James M. Wallace, Vice-President, and Mrs. Margaret H. Rogers, Director, Copy Research Bureau, for their interest and cooperation.

not directly related to the product. Layout D included pictures of miscellaneous content differing from Layouts A, B, and C in that the pictures were more elaborate and contained a much greater amount of detail. For all layouts, the slogans were appropriate to the picture and the product, and were devised in accordance with good advertising practice. The brand name was at the bottom of the advertisement in each case; but for each product, the positions of picture and slogan were alternated, the picture being at the top in half the ads, and the slogan at the top in the remaining half. The twelve ads were photographed in black and white, and 2 × 2 inch glass slides were prepared for each. These ads were supplied by the Copy Department of N. W. Ayer & Son, Inc.

Subjects. The subjects were sixty male undergraduate and graduate students ranging from eighteen to thirty-three years of age.

Apparatus. The GSR apparatus was similar to the alternating current circuit described by Grant (**3**). A Rubicon mirror galvanometer was used and the deflections accompanying changes in skin resistance were read off a scale graduated in centimeters.

Electrodes for the GSR apparatus were dry, polished silver discs attached to the palm and back of the right hand. Constant pressure was maintained by a clamp arrangement which, though comfortable, restricted excess hand movement.

The slides were projected on a daylight screen by a Keystone viewer operated on time exposure.

Procedure. The experiment was conducted in a dimly lighted room so situated that there was a minimum of distraction from the general activity of the building. The subject was comfortably seated approximately ten feet from the screen and so situated that the GSR apparatus and the viewer were not within his line of vision. The size of the image on the screen was approximately three feet square.

The electrodes were attached to the subject's hand and at least five minutes were allowed for adaptation of the subject and stabilization of GSR light deflections. Instructions were given to observe the advertisements as if the subject were glancing through a newspaper or magazine. A sample slide was presented followed by the twelve test slides.

The order of presentation was planned in accordance with the Latin square principle in such a way that over the group of 60 subjects each advertisement occupied a given ordinal position the same number of times. The twelve ads were arranged into five 12 × 12 Latin squares, and each of the 60 subjects was assigned at random to one of the rows of one square. The influence of contiguity upon responses to consecutive ads dealing with the same product or having the same layout was controlled by designing the Latin squares so that the product and layout effects were distributed systematically over the twelve ordinal positions. The first four ads in any row constituted one permutation of the four layouts, the second set of four was another permutation of the layouts, and the third set of four still another permutation. Similarly, the first and each succeeding set of three ads was a permutation of the three products. For example, the sequence presented to one subject was:

Order:	1	2	3	4	5	6	7	8	9	10	11	12
Advertisement:	3C	1B	2D	3A	2B	1D	2A	1C	3D	1A	2C	3B

Here the numbers 1, 2, 3 identify the three products, and the letters A, B, C, D, the four layouts; i.e., "3C" indicates that the first stimulus presented to this subject was the ad representing Product 3 in Layout C. The first group or block of four, comprising orders 1–4, has each of the four layouts represented once; and so do the two remaining blocks, 5–8 and 9–12. Orders 1–3 for this subject include each of the three products, as do orders 4–6, 7–9, and 10–12. The order of presentation was, of course, different for every subject, since 60 permutations of the basic sequence were used, each fulfilling the same conditions concerning the distribution of products and layouts.

The layouts were projected for four seconds, during which time the GSR deflections were recorded. There was an interval of approx-

imately thirty seconds between the presentations of successive stimuli.

Results

Table 1 contains the mean deflections in centimeters for each of the twelve advertisements individually, and the means for the product and layout groups. It will be noted that the mean deflections for the three products vary but little. The means for layouts, however, show considerably greater dispersion; the largest being associated with Layout C, which featured pictures of a young couple (a not unlikely result from a sample of college men), and the smallest with Layout A, the chief characteristic of which was a large picture of the product or service advertised.

As is often the case with data of this kind, distributions of the deflections were positively skewed, and there was a distinct tendency for the means and standard deviations among subjects to vary together. Applications of the usual statistical tests led to the rejection of the assumptions of normality and of homogeneity of variances and to the application of a logarithmic transformation to the data. The resulting distributions were more nearly normal and Bartlett's chi-square test of homogeneity (2, pp. 195–197) indicated that the null hypothesis could not be rejected at the 10% level. This degree of correspondence was considered adequate to warrant the application of analysis of variance.

Table 2 contains the results of the analysis. The residual mean square was used as the denominator in all F-tests. This residual mean square includes, in addition to "error" effects, any second-order interactions which may be present plus a component arising from response differences attributable to the order of presentation, or time sequence. The mean GSR plotted against order of presentation yields a negatively accelerated curve, dropping from a mean of 3.70 cm. on the first presentation to 2.05 cm. on the twelfth and indicating that some sort of adaption was in progress throughout the experiment. As a consequence of the experimental design, this circumstance operates primarily to inflate the residual mean square and has the effect of lowering all F's. It would be possible to make adjustments by taking account of the regression of GSR on order of presentation, but the conclusions drawn from the study would not be affected.

From Table 2 it will be noted that differences among subjects and differences among layouts are significant. Differences among products are, if anything, smaller than one would expect to find by chance, though not significantly so. All three of the first-order interactions are significant. The interaction of products and layouts is interpreted as indicating that, in the sample examined, the GSR response to an advertisement of a given product depends upon the way the advertisement is presented; i.e., that the most effective kind of layout for one product is not most

Table 1

Mean Deflection (in Centimeters) for the Twelve Advertisements Classified by Product and by Layout *

Layouts	Products 1	2	3		Row Means
A	2.73	1.85	2.64		2.41
B	2.28	2.52	3.58		2.79
C	4.20	3.52	2.22		3.31
D	2.66	3.05	2.76		2.82
Column Means	2.97	2.74	2.80	Grand Mean	2.83

* N = 60 for each cell.

Table 2

Analysis of Variance of Logarithmically Transformed GSR Data

Source	Sums of Squares	d.f.	Mean Square	F
Products	.133	2	.066	.74
Layouts	.867	3	.289	3.25*
Subjects	18.880	59	.320	3.60**
Interactions:				
P × L	2.111	6	.352	3.96**
P × S	21.239	118	.180	2.02**
L × S	21.241	177	.120	1.35**
Residual	31.507	354	.089	
Total	95.978	719		

* Significant at 5% level.
** Significant at 1% level.

effective for another. The two interaction effects involving subjects indicate that response tendencies to different layout styles and to different products vary from individual to individual, depending upon differences in needs, experiences, preferences, or other individual characteristics.

The interpretations presented above should not be extended to apply to advertised products in general, to layouts in general, or to individuals in general. Neither the products nor the layout styles used are random samples of any clearly defined "populations" of products or layouts. The fact that the products were given fictitious brand names means, of course, that these findings cannot be generalized to advertisements of familiar brands toward which attitudes and response tendencies may be present. The sample of subjects (college men) is highly restricted, and yields no dependable predictions concerning responses to advertisements by people generally.

A factor whose effects cannot be precisely evaluated in this design is the brightness differences among the projected slides. These differences, which were considerable, were not anticipated, and no provision could be made for them in the factorial analysis. The brightness of each slide was measured with a Weston photometer and compared with the average GSR for that slide. The coefficient of correlation between brightness and average GSR is $-.06$, which indicates little or no relationship between GSR and brightness. Means of the brightness measures grouped according to product and layout do not demonstrate any significant differences. It is concluded that this variable did not unduly influence the experimental findings.

Summary and Conclusion

The desirability of a method for selecting the most effective advertisement from a group of preliminary layouts is recognized by advertisers, and recently some use has been made of the galvanic skin response (GSR) for this purpose. A factorial experiment was designed using the GSR in an attempt to differentiate the reaction-producing characteristics of layouts. Twelve ads were used, comprising four layouts for each of three products. The GSR responses of sixty male subjects to these advertisements revealed significant differences among layout styles and among subjects, but no significant differences among the three products. All first-order interactions of products, layouts, and subjects were significant.

The results of the study demonstrate that the GSR is sensitive to differences in layouts and that further research may lead to the development of practical techniques for testing the effectiveness of advertising copy before the costly publication stage is reached. The study also serves to illustrate the importance of careful planning and design of GSR research, including the use of proper experimental control and efficient statistical methods.

Received August 1, 1950.
Early publication.

References

1. Eckstrand, G., and Gilliland, A. R. The psychogalvanometric method for measuring the effectiveness of advertising. *J. appl. Psychol.*, 1948, **32**, 415–425.
2. Edwards, Allen L. *Experimental designs in psychological research.* New York: Rinehart and Co., 1950.
3. Grant, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149–151.

Book Reviews

Anastasi, A., and Foley, J. P. *Differential psychology* (*Rev. ed.*). New York: Macmillan, 1949. Pp. xv+894. $5.00.

Anastasi's 1937 *Differential psychology* is now enlarged by four new chapters covering the basic concepts of psychological testing, the biological and psychological factors in simple behavioral development, and the effects of schooling on intelligence. Other material, accrued since 1937, on trait organization and socio-economic differences has been added to the earlier discussion and is here reorganized into two substantially new chapters. Eighteen other chapters which conform closely to their former headings and sub-headings, have been re-written to include researches of more recent date. Outstanding inter-disciplinary contributions are certain studies from genetics, anthropology, and sociology.

Sectional additions, scattered throughout the book, comprise: the measurement of special aptitudes, conditions which affect the shape of the distribution curve, the nature of heredity, the nature of environment, the heredity-environment relationship, popular misconceptions regarding heredity and environment, structural and functional characteristics, the concept of unlearned behavior, the study of practice as an approach to the heredity-environment problem, typical findings on the improvement of mental test performance in age, the constancy of the I.Q., the study of family pedigrees, the search for components of physique and temperament, constitutional type as a social stereotype, profile analysis, what intelligence tests measure, a cultural concept of intelligence, evaluation of group differences, the comparative achievement of different races, and the cultural frame of reference in behavior, developmental stages, language, and human nature.

The book is divided to orient the reader historically and methodologically, to present an analysis of individual differences, and then to discuss group differences. Studies from sociology and anthropology are integrated with research in psychology to give a larger perspective on common problems.

This second edition reflects a more studied consideration of content and aptness of expression, and shows growth in unraveling theoretical complexities. The obscured operation of heredity and environment in group differences is analyzed in a more mature fashion, attention being focused on newer insights from recent methodological refinements in the construction and application of psychological tests, from the constantly accruing longitudinal studies, and from the interest expressed in the fundamental nature of intelligence and personality through analyses of their specific components.

The new edition affords a useful compendium for professional workers and students who cannot themselves read, digest, and coordinate original research publications. The main service of this book, however, is in the training it gives in being alert for "common pitfalls and sources of error in interpretation of obtained results," with the authors repeatedly puncturing conclusions drawn from insufficient data, unreliable tests, poorly designed research experimentation, ineffective controls, inadequate statistical treatment, or just plain immaturity of psychological understanding. This second edition is well documented, well illustrated, well arranged, and well written.

Gladys C. Schwesinger
California Youth Authority

Calhoon, Richard P. *Problems in personnel administration.* New York: Harper and Brothers, 1949. Pp. xii+540. $4.00.

The emphasis in this book is on "the understanding of personnel problems: the reasons therefor, with methods of attack and solution so that the feel of actual problems can be more real to the student of personnel administration whether he is in the field or in the classroom." With this objective in mind, Calhoon presents a critical evaluation of present developments in personnel policy and programs.

Through the twenty chapters in the book, Calhoon treats the various kinds of problems faced by personnel administrators, classified in the usual way around selection, training, wage and salary administration, etc. For each phase of a personnel program, Calhoon at-

tempts to show its place in the total program, to point out common errors or weaknesses of present practice, and to make suggestions for a sound program. Practical problems arising in actual administration are stressed throughout. So, for example, in discussing "executive training," Calhoon judges the practice of a one-to-two year rotating program for college graduates as a "dubious procedure at best," points out difficulties such as maintaining motivation and getting the trainee accepted by line supervisors and employees, and then presents the case for having general training follow rather than precede a definite job assignment. Literally hundreds of personnel problems are discussed in this fashion. The book thus becomes a field manual for personnel administrators and a guide for those in training for personnel administration.

As such, its value depends upon the validity of the analyses, criticisms, and suggestions made. Unfortunately, there is inadequate research behind many of the opinions and beliefs upon which personnel policies and practices are now decided. To supplement his own opinions and experience, however, Calhoon reports the results of a questionnaire survey of nearly 600 personnel administrators on nearly all of the major personnel problems. In addition, the NICB surveys and AMA publications are frequently referred to. It can be assumed that, right or wrong, here is a distillation of the best opinion in the field, at least as viewed and reported by one person.

To facilitate its use in personnel training situations, either in school or in industry, Calhoon has ended each chapter with practical discussion problems, projects, and role-playing situations. These study aids are well chosen and, if used properly, will vitalize the use of the book as a text. Its greatest weakness, however, is the lack of reference material since there is no attempt to survey the supporting evidence in the literature and there are no lists of suggested readings. Particularly neglected are references to psychological literature; only one psychological journal article and only a few texts in industrial psychology are referred to and those but briefly.

Albert S. Thompson

Teachers College,
Columbia University

Stuit, Dewey B., Dickson, Gwendolen S., Jordan, Thomas F., and Schloerb, Lester. *Predicting success in professional schools.* Washington, D. C.: American Council on Education, 1949. Pp. v+187. $3.00.

This book summarizes present knowledge gained from experience and research in predicting success in professional schools. It is edited under the sponsorship of the Committee on Student Personnel Work of the American Council on Education and is based on a series of technical bulletins prepared by the Veterans Administration for use by vocational advisers. Professional fields included are engineering, law, medicine, dentistry, music, agriculture, teaching, and nursing.

The already comprehensive compilation of information in the Technical Bulletins has been extended in its usefulness by a very good editing job. A valuable introductory chapter has been added on problems and techniques of prediction. This contains such a good discussion of the problems of prediction, and limitations of research, that it would merit reading and re-reading by all practitioners. Mention is made of the absence of predictive knowledge relative to such factors as interest, motivation, and personality, but this is not, in the writer's opinion, strongly enough emphasized, since the burden of effort is first to establish the importance of what quantitative findings exist. There is, however, a good discussion of the dearth of investigation relative to the influence of environmental factors in the prediction of professional success. The authors wisely caution continuously, not only in this chapter, but also in the concluding statement for each profession, against applying generalizations for prediction to individual cases.

The general discussion of tests as predictive indices omits raising the question of whether tests measure what they purport to measure. It is also somewhat disappointing that although there is emphasis on the restriction of the use of clinical instruments (i.e., personality tests) to professional workers with adequate background in Clinical Psychology, there is not equal caution expressed regarding the interpretation and clinical use of any testing in the counseling process.

For each profession helpful general informa-

tion concerning national requirements precedes the presentation of research findings. The way in which this material is organized for presentation is particularly good. For example, the effect of length of pre-training, native vs. transfer training, etc., are considered separately, followed by an evaluation of each kind of test and combination of indices. A summary and comment on implications for counseling completes each section. The book is in highly readable form, an excellent job having been done in sifting out and simplifying the important findings. There is a useful table of essential qualifications for each field of professional training, with a suggested battery of tests. These test recommendations are open to question, especially recommendations given in the personality area. In addition, no mention is made of the fact that, in generalizing regarding the lack of predictive value of interest inventories, such as the Strong, only a single key has been used, when it is generally considered that patterning or constellation is more diagnostically significant.

Despite these few omissions, this book is an invaluable compilation for any high school or college counselor. It would seem also that in carefully selected cases it might profitably be made available to counselees who are considering one or another of the professional fields covered. In any case, filling a major gap, it is an imperative addition to the working materials of the professional counselor.

(Mrs.) Barbara A. Kirk

Counseling Center,
University of California,
Berkeley, California

Schramm, Wilbur, Editor. *Mass communications.* Urbana: University of Illinois Press, 1949. Pp. xi, 552. $4.50.

Mass communications is a collection of readings. It claims to be nothing beyond that, and the sub-title clearly states that it is a book of readings. Consequently, there is no reason to expect that prospective purchasers will be misled into expecting anything else.

There are several indications that this collection of readings fills a real need, rather than just being another book. Experience with the symposium, *Communications in modern society*, of the Institute of Communications Research of the University of Illinois suggested the need for something that would be broader in scope. Furthermore, there were a number of more recent publications that deserved consideration; so the Director of the Institute selected and edited this collection of readings. Since all of the selections have been used in the Institute, this publication rests on a background of actual experience.

In a field as broad as this one, anyone who attempts to select and edit a limited number of readings obviously is confronted by several very difficult problems. His success in solving these problems largely determines the value of the publication.

Providing a satisfactory degree of continuity is one of these problems. Schramm's solution consists mainly of grouping the readings and giving a short introduction for each section. The section on the development of mass communications includes readings on the history of the newspaper, motion pictures, radio and international communications. The control and support of mass communications are considered in a second section. The third section consists of five readings on various aspects of the communications process from the standpoints of general semantics, psychology, and anthropology. The three remaining sections cover content, audiences, and effects.

The appendix presents several pages of a wide variety of facts and figures.

Judged by conventional standards used to evaluate most publications, this collection of readings would have to be rated as spotty in coverage and below par in continuity. However, it should not be judged by such standards. The gaps are the result of the characteristics of this field and are beyond the control of any editor. The raw material for an integrated treatment of the subject just does not exist.

Perhaps an even more serious problem is the headache of deciding how much space to give each paper. Complete reproduction of each article and chapter selected would be possible only by limiting coverage to a point at which the objective would not be accomplished.

The longest reading is the summary of Borden's *The economic effects of advertising*, which is 41 pages. Second in length is Llewellyn White's *Ragtime to riches* (28 pages), which

is one of the two readings by the same writer on the development of American radio. Only two other readings are longer than 20 pages. Twenty-three readings run ten to twenty pages. Fifteen readings, not counting the section introductions, are shorter than ten pages.

Inclusion of the shortest readings raises a question. For example, the shortest reading is one page of definitions of the devices of propaganda from *The fine art of propaganda.* Another reading consists of three tables showing changes in heroes, settings, and themes in large-magazine fiction (from Johns-Heine and Gerth). The summary of Kenneth Baker's study of radio programming takes up only two pages, and so does the presentation of the Flesch readability formula.

At first glance these short readings appear to contribute to the impression of inadequate continuity. In addition, such treatment of a subject might be misleading. However, including them does increase the coverage of the volume as a whole. Since the book probably will be used most often in class situations that permit further explanation and since individual users outside of classes probably will consult the original sources, the reviewer agrees with Schramm's decision to include them.

When complete reproduction is not practical, one is faced with a problem of what to omit. In several such cases, Schramm reports results and conclusions at the expense of description of the methods. Although this may tend to cause some concern among technicians, it is the only practical approach which is in line with his objectives. Obviously, this book was not intended for research technicians in a strict sense. Many of the readings are what tough-minded research people probably would consider "essays," although some of them are relatively quantitative. This is as it should be. Any other approach would present a distorted picture of the broad field of mass communications.

The usefulness of this book depends on the need that any particular person might have for a collection of readings on this subject. It provides convenient material for class purposes, and it at least minimizes the amount of digging necessary to get an introductory overall picture of this field. There is ample evidence of both good judgment and experience in the selection and editing. Any deficiencies result from the nature of the available material and the weaknesses inherent in any publication which is based upon the selection and editing of previously published manuscripts.

Alfred C. Welch

Knox Reeves Advertising, Inc., Minneapolis

Massing, Paul. *Rehearsal for destruction.* New York: Harper and Brothers, 1949. Pp. 341. $4.00.

The book is well-written, carefully documented, and scholarly in style, but is strictly a "who did what" (with dates) historical analysis. The absence of any systematic discussion of the group or individual psychological determinants and factors in German anti-Semitism in the period covered (1871–1914) leaves little of interest except for the most zealous student of the subject. The broader purposes of the study may well have required this painstaking chronicling approach, but the result is a volume of only slight interest to most psychologists.

Harrison G. Gough

University of California, Berkeley, California

McCord, Carey P., and Witheridge, William N. *Odors: Physiology and control.* New York: McGraw-Hill, 1949. Pp. 405. $6.50.

In these days of ever-increasing specialization it may appear as temerity to attempt covering in one volume most of what is known about odors—from the anatomy of the olfactory system and the physiology of olfaction to the legal aspects of odor nuisances; from the theoretical considerations of the mode of action of the chemical stimuli to presentation of practical procedures for odor cancellation and counteraction; from classification of the odors to the use of odors as warning agents in fuel gases and as a method of broadcasting alarm in mining emergencies; from discussion of the rather unsuccessful efforts at clarification of the relationship between chemical structure and the quality of odors to the control of some industrial odors by altering the chemistry of the manufacturing processes; from odor of the human body, in health and disease, to household odors and odors of water and foods. The

authors were concerned, primarily, with the offensive odors and the engineering of odor control. The applied psychologist will read with some interest the chapters on industrial odors and the offensive trades. He might be consulted, on rare occasions, about making an odor survey; he will find in the present volume valuable practical suggestions.

Because the area covered by the book is so large, many topics are treated rather sketchily. This appears to be particularly true of matters psychological. Thus, in connection with the question of pleasantness and unpleasantness of odors, mentioned only very briefly, the authors feel that "Such matters so invade the realm of emotions and habituations that further delineation must be delegated to the psychologists" (p. 28). The short section on "mental perturbation" resulting from offensive odors (p. 71) is little more than a sample of "literary psychology." The experimental psychologist will note the new electrochemical theory of odor stimulation (p. 22), based on Linus Pauling's ideas about molecular architecture and biological reactions in general and odor sensations in particular, but he will be aware of its highly speculative nature.

There is a large and moderately useful bibliography (124 pp., some 3,500 references), arranged by chapters. The authors stated that several hundred of the items were cited from secondhand sources. It would have been useful to identify this material by some simple symbol and to separate it from references actually seen.

Josef Brožek

University of Minnesota

Bellows, Roger M. and Rush, Carl H., Jr. *Workbook in personnel methods.* Dubuque, Iowa: Wm. C. Brown Co., 1949. Pp. 102. $2.10.

This workbook is designed for use in an introductory course in personnel methods to accompany Bellows' "Psychology of Personnel in Business and Industry." It attempts to "enable the personnel methods trainee actually to participate in personnel methods problems." It emphasizes the objective approach to problems and gives practice in the handling of quantitative data and in quantitative approaches to human relations problems.

In line with this orientation, the authors have prepared exercises which require the analysis and interpretation of data such as might be obtained in an actual situation and which illustrate typical personnel techniques. So, for example, an exercise on the reliability of criteria deals with inter-rater agreement, one on recruitment requires computing average test scores of applicants from three recruitment sources, one on selection involves an item analysis of application blank items, one on training requires plotting learning curves on two groups of industrial trainees, etc. Practice is given in the use of Taylor-Russell tables, an interview guide, an item discrimination nomograph, norms, scattergrams, a merit rating form, a sociogram, the Flesch readability formula. Altogether 22 exercises related to 14 of the 20 chapters in the text are provided.

In general, the exercises are well worked out; the instructions are clear, the data are well presented, the interpretation questions are thought-provoking. The common danger of a workbook's becoming merely a study guide to the text has been avoided. It is not the "case-study" type of textbook supplement but one which is designed to give the student an exposure to some of the simpler tools used in the scientific approach to personnel problems.

Albert S. Thompson

Teachers College,
Columbia University

Pressey, Sidney L. *Educational acceleration: appraisals and basic problems.* Columbus, Ohio, Ohio State University: Bureau of Educational Research, 1949. Bureau of Educational Research Monographs, No. 31. Pp. xiv + 154. $2.50, paper. $3.00, cloth.

This book, which should be of interest to all who are concerned with policies in all levels of educational institutions, presents the results of a research project conducted at Ohio State University. According to the author, "The effort was made broadly to investigate a problem, experimentally to try various means for dealing with it, and systematically to appraise the results." In this, Pressey has produced a stimulating and valuable piece of work.

Until the last war resulted in pressure for educational acceleration, discussion of it and attempts to try it were largely characterized

by *a priori* thinking and little data. The changes brought about by the war, however, provided an opportunity to study the effects of acceleration on large numbers of students. The author did this extensively and considered the problems carefully.

Pressey has shifted the emphasis to the question of what justification there might be for continuing to prolong adolescent status and delay mature social functioning and productivity. He has found little or none. The work is organized around three basic concepts which lend clarity and direction to the studies: maturity, prime (defined as peak biological potentiality), and individual differences.

After chapters dealing with these basic concepts and with the opinions and studies of other educators and psychologists, the rest of the book seems to fall into three main sections: one part deals with the trend toward delayed completion of training; another presents the data bearing on the results of acceleration; the third is concerned with alterative methods of acceleration.

In the latter two sections, which are the more important, Pressey shows that, contrary to popular notions, accelerated students tend not only to be superior in academic achievement (even when matched with non-accelerates with equal aptitude), but they take part in as many social activities, and, after graduation, make equally good or better job and life adjustments. Included in this section is a study demonstrating the need for *guidance* in selecting students for acceleration.

The major alternative method of acceleration (in addition to summer attendance and heavier course loads) suggested is credit by examination. This is shown to have no adverse effects on later scholarship. In addition, data are presented, taken from questionnaires and interviews, which show that in many cases, the accelerates feel that the present most common methods place them under undue strain. The conclusion drawn is that before even guided acceleration can be effective, educational institutions will have to revise their rules and mores to avoid making the accelerates feel out of place or too pressed.

The most serious criticism of this book lies in the area of the presentation of the data. The conclusions are based on differences between percentages of various groups meeting certain criteria. Yet nowhere is there presented a test of the significance of these differences. Pressey says, for example, "The differences are so small as to be of questionable significance, each taken by itself. Taken together, they show a consistency . . . ." While such consistency may argue for the acceptance of the results, still professional readers will probably have as hard a time as the reviewer in understanding why so eminently able a person as Pressey should have left this crucial step in the analysis undone or unreported.

Pressey concludes the book by reemphasizing the need for guidance in any flexible program; he then makes eight specific recommendations which, he feels, might serve as guideposts to people willing to come to terms with the three basic concepts which he presented at the beginning.

Whatever the shortcomings of the analysis, Pressey has done a masterful piece of work in clearing the ground so that the real issues can be clearly seen; other investigators may well be able to explore the problems further and more adequately because of his work.

John W. Gustad

Vanderbilt University

Weston, H. C. *Sight, light and efficiency.* London: H. K. Lewis and Co., Ltd., 1949. Pp. xiv + 308. 42s.

This English treatise, based largely upon the results of recent research sponsored by the Medical Research Council and the Department of Scientific and Industrial Research in Great Britain, is concerned mainly with illumination in relation to industrial and occupational efficiency. After preliminary attention to the sense of sight and causes and symptoms of eyestrain, consideration is given to occupational demands upon sight, the facilitation of visual tasks, lighting and visual efficiency, incentive luminance and color, testing vision, and protection of sight at work. Recommended levels of illumination are listed in the appendix. There is an abundance of appropriate figures, tables and illustrations.

There are many contributions in this book which may be cited with approval. Among these are the following: (1) The industrial psychologist, the illuminating engineer, and

others interested in visual efficiency will find material of practical value. (2) To a large degree the cited experimental data provide a sound basis for the conclusions presented. (3) One should know what acuity is needed for comfortable work rather than acuity just sufficient to permit the job to be done. (4) The extent to which contrast and size are interdependent factors affecting the severity of the visual task is clearly explained. (5) It is stated (page 170) that "however high illumination is made, objects which are not very nearly alike either in apparent size or contrast, or in both these characteristics, never become equally easy to see." These findings should help to correct the views of those American writers who claim that the effects of small size and poor contrast can be completely compensated for by increases in illumination. (6) Methods of illuminating work objects are effectively described. (7) The psychological feeling of well-being that comes with adequate illumination combined with desirable combinations of color and brightness of painting receives a desirable emphasis. (8) For the most part, the schedule of recommended values of illumination are adequate (not excessive) in terms of experimental findings.

Careful evaluation will raise questions on certain specific details. A few of these criticisms follow: (1) During recent years it has often seemed practical to compute the illumination level needed for a specific visual task. A job analysis reveals size of detail and brightness contrast. Then by reference to laboratory studies of visual acuity in relation to size of detail, contrast and illumination, the level of illumination for the job is computed. Weston has worked out nomograms to facilitate such determinations. Although this procedure is helpful at the present stage of our knowledge, it is a dubious practice. It is questionable whether it is valid to make a direct transfer from visual acuity data to a relatively complex seeing situation where integrated visual patterns are involved. Furthermore, recommendations which rest on estimates from threshold measures have not in most cases been validated for the actual visual work activities. When this has been done, the computed illumination values have been found to be greater than needed. (2) When the author cites the need of 500 foot candles or more for lace-making because workers prefer to work out in the sun rather than in poorly illuminated interiors (page 153), his inference is unwarranted. There is no other choice, such as a well lighted interior, available. (3) It is possible that other factors than differences in illumination (page 162) affect the shape of the work curves in winter as compared with summer. (4) The tentative inference (page 142) that white print on black background is more readable than black on white is not supported by experimental findings.

Industrial psychologists and illumination engineers should welcome this well organized and well written book on the relation between illumination and visual efficiency. Careful reading should reveal how extravagant some of the illumination intensities recommended in this country are.

Miles A. Tinker

University of Minnesota

Ward, Roswell. *Out-of-school vocational guidance.* New York: Harper and Bros. 1949. Pp. xiv + 155. $2.50.

Admittedly, the literature of vocational guidance dealing specifically with the "out-of-school" population is limited. The author draws upon extended experience as a vocational counselor and as an organizer of out-of-school vocational guidance services in many states, in an attempt to fill the gap.

"Vocational guidance" is held to be a fact-finding, coordinating, planning and referral service. The "vocational counselor" who is the chief figure in this service, does more than counseling (the "planning" part of vocational guidance). His effectiveness is determined most by his skill as a fact-finder and his ability to coordinate and make use of information and resources.

The narrowing of the traditional definition of "vocational guidance" shows its value both in the organization and in the operation of a vocational guidance service. Many practical ideas are presented for carrying on an organized community service to meet the continuing need for vocational guidance.

The author does not hesitate to present rather dogmatically certain ideas which his

own experience has convinced him are valuable. Among these, the following are typical:

The identification of the "non-selective" worker permits economies through the use of less intensive forms of counseling.

Adherence to a classification of "vocational problems" according to symptoms is consistent with the defined function of the vocational counselor as a maker of plans for vocational adjustment and as an "action-centered" rather than "applicant-centered" fact-finder and referral agent. The categories of this classification (each preceded by the word "Vocational") are: Immaturity, Confusion, Insecurity, Misdirection, Fixation, Conflict.

A sharp distinction between "interest" and "motivation" has implications for the fact-finding activities of vocational counseling. An eighteen-item classification of vocational motivations is presented.

An insistence that vocational counseling must be tied basically to vocational information and vocational economics rather than to psychology, has implications for the selection and training of vocational counselors. Prejudice is revealed here which is carried to the point that the training of vocational counselors would be taken out of schools of education and departments of psychology. The author likens the contrast between vocational guidance as an adjunct of education and as an out-of-school service, to the difference between fire prevention and fire fighting.

A criticism of "mirror counseling" is a further revelation of the same prejudice. Consistent with the fear of "mirror counseling" is a distrust of a psychologist's ability to become a vocational counselor. The psychologist should be only a resource person available to the "applicant" (not "client") by referral from the counselor.

One of the limitations (other than the fairly obvious effects of the revealed prejudices) of the volume lies in the heavy dependence upon the "applicant" either in the interview or by filled-in forms, for the data about himself essential to his vocational planning. The usual categories of data are mentioned but techniques for securing and evaluating them are left largely to the judgment, attained chiefly through experience, of the vocational counselor.

It is the opinion of the reviewer that when the author's prejudices are fully discounted, the book stands as a practical challenge to many traditional practices. Also, the book should be thought-provoking, rather than offensive, to psychologists and educators concerned with carrying on counseling services and with the preparation of counselors.

Fred M. Fowler

Department of Public Instruction, The State of Utah

Allport, G. W. *The individual and his religion.* New York: The Macmillan Company, 1950. Pp, x + 147. $2.50.

Almost independently of its merits, this book is important in three respects. First, and I think of greatest significance, is the fact that an American psychologist of first rank and theoretical interests has addressed himself to this topic. As Allport points out, contemporary psychologists avoid the subject of religion about as much as James and his contemporaries avoided any detailed treatment of sex. (Sexuality, by the way, gets strikingly little attention in Allport's discussion of the organic "origins of the religious quest," pages 9–12.) The state of mind of modern Western man, including many of the intellectuals, is such that even the most atheistic psychologist, if interested in personality, ought to concern himself with the religious impulses in somewhat more detail than to dismiss them airily as "sublimations," "failure of nerve," and the like. That a writer of Allport's stature has done so may stimulate others, who do not share his scientific views about personality, to do the same. Secondly, this book will be useful to counselors as bibliotherapy for clients experiencing religious conflict, especially in the case of students whose conflicts arise via contact with psychology and allied sciences. Thirdly—and here I part company with Allport—the book is a challenge to the determined and wholehearted secularist in psychology, because it is quite obiously *not* a disinterested psychological causal analysis but includes a not-too-subtle bit of religious propaganda, the author's avowal (p. vii) to the contrary notwithstanding. I do not know the facts about Allport's own religious convictions, if any; but unless I am mistaken he has a

religious urge two sigmas above the American psychologists' mean. Let me cite a few random examples having this flavor. "The universe is simply incomprehensible. Fragments of it may be fairly well understood, but certainly not the design of the whole" (p. 18). "The fact . . . is that scientific thought is known . . . to be able to cover only part of the ground. It goes a long way but not far enough" (p. 20). ". . . even in this age of technology and social disintegration when skepticism hangs heavily upon the horizon. . ." (p. 39).

Certain purely "psychological" theses are presented without the qualifications one would like to see in a book that will get into the hands of non-psychologists. On page 63, we find that ". . . in only slight degree, if at all, is this energy drawn from the reservoir of organic drives"—treating Allport's important and respectable theory of autonomy as an established principle of psychological science. We are assured (p. 65) that if a strong religious sentiment fades away, ". . . we can be certain that religion was never a central feature of the personality" (Allport's form of Calvin's doctrine of perseverance, perhaps?).

The book is marred by a few minor mistakes, such as Allport's canonizing of Thomas à Kempis (p. 124), and the lugging in of that dreadful old chestnut, so often refuted by careful logical analysis, about determinism's incompatibility with praise, exhortation, and effort (p. 115). The presentation of the positivist critique under "referential doubting" (pp. 117–121), especially the end of the paragraph at the top of page 119, is incredibly superficial. Positivists do not "condemn" poetry, for it does not claim to assert facts. "Legal discourse" is primarily a metalinguistic examination of "what entails what," and is, so far as I know, admitted by most positivists. The lumping together of "poetic, artistic, legal" discourse as meaningless to positivists shows a carelessness which in a less respected writer would almost suggest a lack of intellectual integrity.

Finally, the book fails to satisfy because of a deep-seated methodological mistake. It assumes that a really well-rounded psychological discussion of religion *can* be given without any concern with the truth-character of religious beliefs. By this device, Allport tries to have his cake and eat it too; he grinds the religious axe and at the same time avoids the necessity of defending any particular religious thesis. But certain of his "psychological" assertions can simply not be made unless the prior question of validity has been settled. For example, he several times rejects the common characterization of religion as "wish-fulfilling." But he is at great pains to exhibit its role in satisfying needs. Now, if the objects of (propositional!) religious belief do not exist, or if the evidence for their existence is slight, those who have such beliefs, and have them on the psychological bases Allport describes, are surely engaging in wish-fulfilling, whether Allport likes the phrase or not! The constant use of evaluative terms and remarks shows that Allport is not carrying out a *purely* causal analysis, but he never comes to grips, except in a very sketchy and dilettante manner, with the cognitive issues *about* religion which are involved. The book suffers here from that subjectivism, that making man the measure of all things (I am sure Allport would deny this!) which Bertrand Russell has described as the "characteristic madness of our times." For this reason, the philosophically sophisticated clinician could use the book therapeutically only by permitting himself an essentially cynical view of the religiously conflicted student—"it will help him, because he won't see through it."

Paul E. Meehl

University of Minnesota

Tinkelman, Sherman. *Difficulty prediction of test items.* New York: Bureau of Publications, Teachers College, Columbia University, 1947. Pp. 55. $1.85.

Using a civil service examination for patrolman consisting of 100 5-option multiple-choice items covering arithmetic reasoning, civic information, logical inference, reading comprehension, judgment, vocabulary and other topics, the author had thirty judges rate the difficulty of each item. This rated difficulty (per cent of patrolman candidates expected to answer correctly) was compared with the actual per cent of patrolman candidates passing. The judges "seemed to constitute a professionally competent group of exami-

ners" with from three to thirty years of experience with the examining agency. The validity of the judges' estimates of difficulty (correlation between estimated and actual per cent correct) ranged from .23 to .77 with a median of .53.

The validity of the pooled judgment of the thirty judges was .76. Nine graduate students in an advanced test construction course, (presumably without three to thirty years experience) repeated the estimates for an entirely different test. For this trained, but inexperienced group (on a different test, to be sure), the validities ranged from .17 to .72 and the validity of the pooled judgment (9 judges) was .81, higher than the value of .76 obtained from 30 experienced personnel examiners.

Analysis of variance techniques were used to identify factors affecting the predictability of item difficulty. Predictability was unrelated to item content—difficulty of vocabulary items was no more accurately predicted than was that of arithmetical reasoning and the differences among content categories could have arisen by chance. Judges were relatively consistent in their ability (or inability) to estimate or guess the difficulty of the items as measured by their performances on odd and even items.

Even within the limits of "validity" represented by a median coefficient of .53, there were large constant errors of over- or underestimation. For only nine of the thirty judges were these constant errors statistically insignificant. Since the constant errors were in both directions, the pooled judgment was substantially accurate.

The judges tended to overestimate the percentage of passes on difficult items and to underestimate it for easy items. The judgments thus exhibited the usual regression toward the mean. There were no significant differences in predictability between "judgment" and "information" questions or between items with high and those with low item-test coefficients (erroneously termed validity coefficients).

The problem is introduced in a civil service setting—and a rather conservatively conceived one. The necessity for basing the selection of items on subjectively judged estimates of difficulty is mentioned on page 1, but *nowhere is the need for estimates of validity—subjective or statistical—alluded to.* The impossibility of experimental tryout of items is ruled out because of the need for secrecy although progressive merit systems have found ways of tryout without violating security.

The tone of the conclusions and implications seems unwarrantedly optimistic. "The results of this study indicate that judges are probably able to estimate relative item difficulty accurately." This conclusion does not seem warranted for a median validity correlation of .53. Moreover, although the "best" three judges have a combined validity of .85, the author fails to note that this is a "back-validated" coefficient, comparable to that computed for a test on the item analysis sample after the best items have been selected by item analysis.

Although Tinkelman's book does shed light on the process of judging difficulty, it will not convince the careful examiner that estimated difficulty is an adequate substitute for empirical data.

Charles I. Mosier

Adjutant General's Office, Washington, D. C.

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

Projective psychology. Lawrence E. Abt and Leopold Bellak, Editors. New York: Alfred A. Knopf, Inc., 1950. Pp. 485. $6.00.

Converting a veterans guidance center. George D. Baruhal. California: Stanford University Press, 1950. Pp. 99. $1.50.

Beginning experimental psychology. S. Howard Bartley. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 483. $4.00.

Problem-solving processes of college students. Benjamin S. Bloom and Lois J. Broder. Chicago: University of Chicago Press, 1950. Pp. 109. $2.75.

Readings in modern methods of counseling. A. H. Brayfield, Editor. New York: Appleton-Century-Crofts, Inc., 1950. Pp. 526. $5.00.

Sex questions and answers. Fred Brown and Rudolf T. Kempton. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 256. $2.95.

Scientific research: its administration and organization. George P. Bush and Lowell H. Hattery, Editors. Washington, D. C.: American University Press, 1950. Pp. 190. $3.25.

Personality, a systematic, theoretical and factual study. Raymond B. Cattell. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 689. $5.50.

The application of measurement to health and physical education. Second edition. H. Harrison Clarke. New York: Prentice-Hall, Inc., 1950. Pp. 455.

Recent experiments in psychology. Second edition. Leland W. Crafts, Theodore C. Schneirla, Elsa E. Robinson, and Ralph W. Gilbert. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 503. $3.50.

Some theory of sampling. William E. Deming. New York: John Wiley and Sons, Inc., 1950. Pp. 602. $9.00.

Psychology. Second edition. Dockeray and Lane. New York: Prentice-Hall, Inc., 1950. Pp. 576. $3.75.

America begins. Richard M. Dorson, Editor. New York: Pantheon Books, Inc., 1950. Pp. 438. $4.50.

The counseling interview. Clifford E. Erickson. New York: Prentice-Hall, Inc., 1950. Pp. 174. $1.75.

Communicating ideas to the public. Stephen E. Fitzgerald. New York: Funk and Wagnalls Co., 1950. Pp. 292. $3.50.

The principles of scientific research. Paul Freedman. Washington, D. C.: Public Affairs Press, 1950. Pp. 222. $3.25.

The psychology of dictatorship. G. M. Gilbert. New York: Ronald Press Co., 1950. Pp. 327. $4.00.

Outside readings in psychology. E. L. Hartley, H. G. Birch, and R. E. Hartley, Editors. New York: Thomas Y. Crowell Co., 1950. Pp. 880. $2.75.

Clinical applications of suggestion and hypnosis. William T. Heron. Springfield: Charles C Thomas, Publisher, 1950. Pp. 125. $3.00.

Textbook of abnormal psychology. Revised edition. Carney Landis and Marjorie M. Bolles. New York: The Macmillan Co., 1950. Pp. 634. $5.00.

Psychiatry for social workers. Second edition. Lawson G. Lowrey. New York: Columbia University Press, 1950. Pp. 385. $4.50.

In my mind's eye. Frederick Marion. New York: E. P. Dutton and Co., Inc., 1950. Pp. 315. $3.75.

Carbon dioxide therapy, a neurophysiological treatment of nervous disorders. L. J. Meduna. Springfield: Charles C Thomas, Publisher, 1950. Pp. 248. $5.00.

Experiments in social process. James Grier Miller, Editor. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 201. $3.00.

Physiological psychology. Second edition. Clifford T. Morgan and Eliot Stellar. New York: McGraw-Hill Book Co., Inc., 1950. Pp. 572. $5.00.

Psychology and art of the blind. G. Revesz. New York: Longmans, Green and Co., Inc., 1950. Pp. 338. $8.50.

The development of a needs and problems inventory for high-school youth. Benjamin Shimberg. Lafayette, Ind.: Studies in Higher Education LXXII, The Division of Educational Reference, Purdue University, 1950. Pp. 78.

Delinquency and human nature. D. H. Stott. Fife, Scotland: Carnegie United Kingdom Trust, Dunfermline, 1950. Pp. 460. $1.00.

Improving management communication. General management series No. 145. New York: American Management Association, 1950. Pp. 26. $.75.

Practical operating problems in personnel administration. Personnel series No. 129. New York: American Management Association, 1949. Pp. 28. $.75.

Industrial applications of medicine and psychiatry. Personnel series No. 130. New York: American Management Association, 1949. Pp. 31. $.75.

German aviation medicine, World War II. Vol. I and II. Prepared under auspices of the Surgeon General, U.S.A.F. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1950. Pp. 1302. $8.50, set.

Wages, employment and personnel problems in a changing economy. Production series No. 187. New York: American Management Association, 1949. Pp. 71. $1.25.

PSYCHOLOGICAL MONOGRAPH

#305
1949

THE DEVELOPMENT AND VALIDATION OF A SET OF MUSICAL ABILITY TESTS

BY

ROBERT W. LUNDIN

$1.00

Dr. Lundin's monograph, #305 of the Psychological Series, is one of the most popular of those recently published.

American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.

McGraw-Hill Books

ADVERTISING PSYCHOLOGY AND RESEARCH. An Introduction

By Darrell B. Lucas, New York University, and Steuart H. Britt, McCann-Erickson, Inc. Ready in December

An authoritative text for students of psychology and advertising. Both authors are psychologists who are working in the field of advertising. They have drawn upon both professional and practical knowledge to combine psychological principles with extensive applications of practical research methods.

MANAGEMENT OF PERSONNEL AND LABOR RELATIONS. New 2nd edition

By Gordon S. Watkins, University of California, Riverside; Paul A. Dodd, University of California, Los Angeles; Wayne L. McNaughton, University of California, Los Angeles; and Paul Prasow, University of Southern California. 974 pages, $5.75

A thorough revision and rewriting of this well-known college text, the second edition, while retaining its original arrangement in general outline, incorporates the many developments which have occurred in the field since its first publication. A great deal of new illustrative material has been added, and its arrangement and logical pattern of presentation render it exceptionally teachable.

PROBLEMS IN LABOR RELATIONS

By Benjamin M. Selekman, Graduate School of Business Administration, Harvard University; Sylvia K. Selekman; and Stephen H. Fuller, Graduate School of Business Administration, Harvard University. *Harvard Problem Books.* 672 pages, $6.00

Maintaining that collective bargaining must be studied as human behavior, the authors have concentrated on the daily "job mix" by which union men and management representatives adapt to legal regulation, to economic factors, and to their institutional and individual needs. The text is suitable for text or supplementary use in courses in labor relations.

PERSONALITY AND PSYCHOTHERAPY. An Analysis in Terms of Learning, Thinking and Culture

By John Dollard and Neal E. Miller, Yale University. *McGraw-Hill Publications in Psychology.* 479 pages (textbook edition) $5.00

This distinctive new text deals with abnormal behavior and psychotherapy, since they bring out clearly factors that are crucial in the dynamics of the normal personality. A wide range of material is integrated: basic principles of learning, socially learned drives, new experimental studies of fear, the dynamics of conflict, emotional problems, cultural conditions of learning, etc. Three great traditions are brought together—psychoanalysis, the exact methods of the experimentalists, and modern sociology and anthropology.

Send for copies on approval

McGraw-Hill Book Company, Inc.

330 West 42nd Street New York 18, N. Y.

JOURNAL OF APPLIED PSYCHOLOGY

VOL. 35

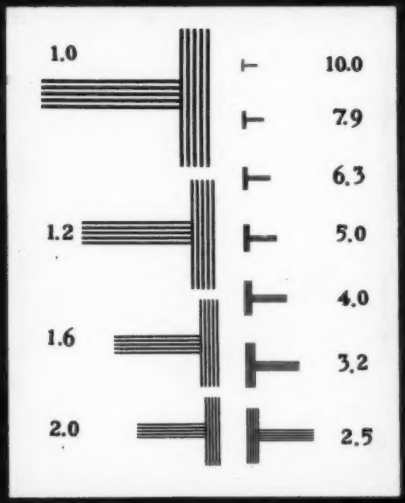
NOS. 1 – 6

1951

PUB. 277

UNIVERSITY MICROFILMS
ANN ARBOR, MICHIGAN 1952

RESOLUTION CHART



100 MILLIMETERS

INSTRUCTIONS Resolution is expressed in terms of the lines per millimeter recorded by a particular film under specified conditions. Numerals in chart indicate the number of lines per millimeter in adjacent "T-shaped" groupings.

In microfilming, it is necessary to determine the reduction ratio and multiply the number of lines in the chart by this value to find the number of lines recorded by the film. As an aid in determining the reduction ratio, the line above is 100 millimeters in length. Measuring this line in the film image and dividing the length into 100 gives the reduction ratio. Example: the line is 20 mm. long in the film image, and $100/20 = 5$.

Examine "T-shaped" line groupings in the film with microscope, and note the number adjacent to finest lines recorded sharply and distinctly. Multiply this number by the reduction factor to obtain resolving power in lines per millimeter. Example: 7.9 group of lines is clearly recorded while lines in the 10.0 group are not distinctly separated. Reduction ratio is 5, and $7.9 \times 5 = 39.5$ lines per millimeter recorded satisfactorily. $10.0 \times 5 = 50$ lines per millimeter which are not recorded satisfactorily. Under the particular conditions, maximum resolution is between 39.5 and 50 lines per millimeter.

Resolution, as measured on the film, is a test of the entire photographic system, including lens, exposure, processing, and other factors. These rarely utilize maximum resolution of the film. Vibrations during exposure, lack of critical focus, and exposures yielding very dense negatives are to be avoided.

THIS PUBLICATION IS REPRO-
DUCED BY AGREEMENT WITH THE
COPYRIGHT OWNER. EXTENSIVE
DUPLICATION OR RESALE WITH-
OUT PERMISSION IS PROHIBITED.